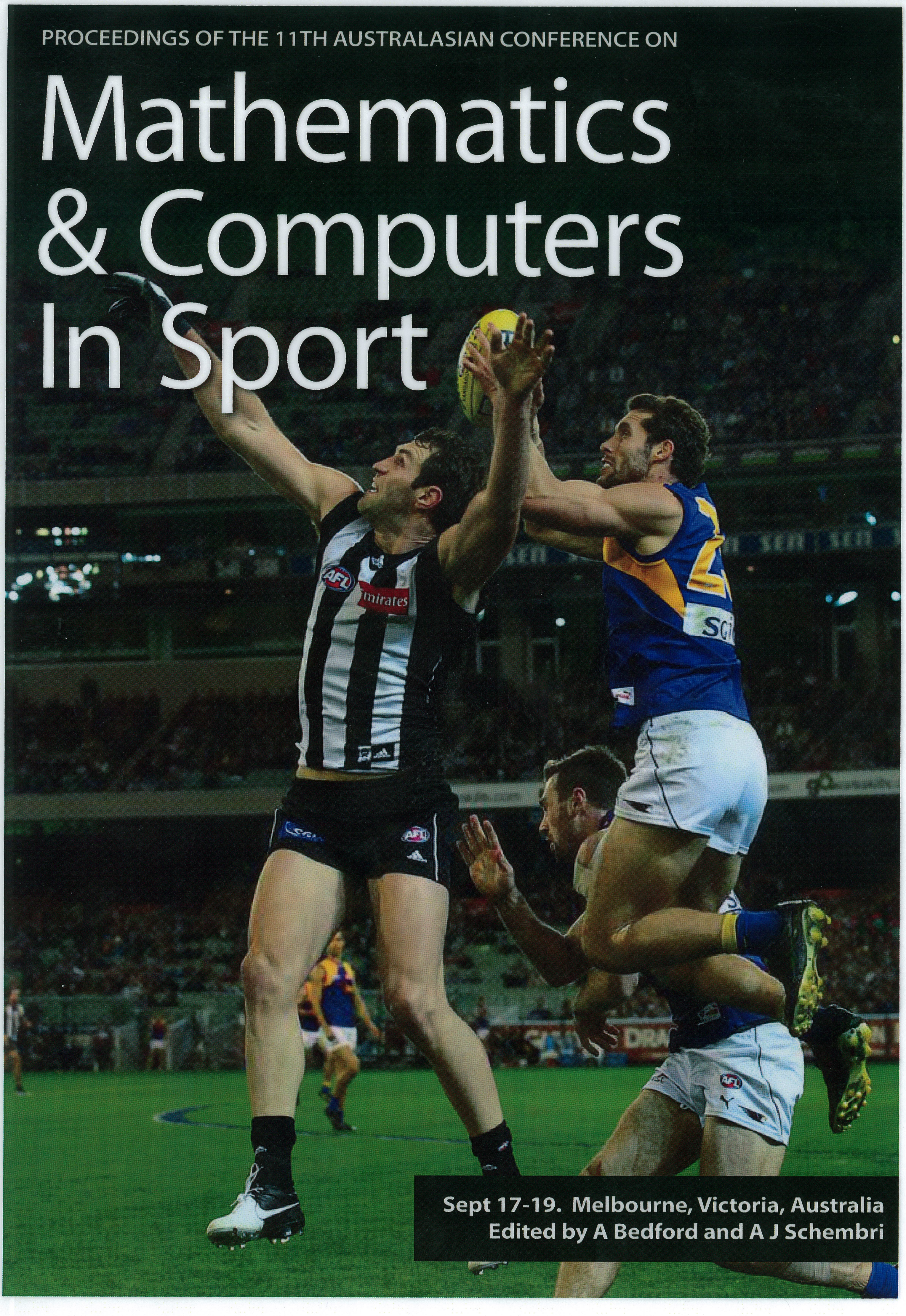
PROCEEDINGS OF THE 11TH AUSTRALASIAN CONFERENCE ON

Mathematics & Computers In Sport



Sept 17-19. Melbourne, Victoria, Australia

Edited by A Bedford and A J Schembri

PROCEEDINGS OF THE ELEVENTH AUSTRALASIAN CONFERENCE

ON

MATHEMATICS AND COMPUTERS IN SPORT

edited by

Anthony Bedford and Adrian J Schembri

11M&CS

September 17th to 19th 2012

Melbourne, Australia

Published by MathSport (ANZIAM)

ISBN: 978-0-9578623-6-4

The papers in these proceedings have gone through a full peer review refereeing process.

Cover Design: Yi Ling Yap

CONFERENCE DIRECTOR'S REPORT

For the first time in its history, the Australasian Conference for Mathematics and Computers in Sport is held in Melbourne. We are very proud of our cosmopolitan city and hope that the visitors to Melbourne will enjoy a good mix of our cuisine and cultures while you are here. The city hosts a wonderful mix of sporting events and venues with the AFL grand final less than two weeks away, so it is very appropriate that the conference is held in Melbourne at this time. Melbournians consider the city to be the sporting capital of world, and currently enjoys number one status as the world's most liveable city.

The conference program includes 36 paper presentations, covering a good mix of sports and methodologies. There are six guest speakers and three panel sessions involving tennis technologies, the Olympic experience and wagering in sport. In addition, something that would only be possible in Melbourne, the AFL Research Committee will present a keynote address. We are very grateful to all the guest speakers for sharing their experiences and their inside knowledge with us.

All full papers in these proceedings have been peer refereed, and we thank all the reviewers for their swift returns given the tight time frame. For assistance in the organisation of 11M&CS, we are very grateful to co-editor Dr Adrian Schembri (RMIT University) who has contributed a significant amount of effort in both compiling and reviewing papers, along with our outstanding scientific committee members. We also extend a big thanks to conference secretary Jaclyn Yap (RMIT University) for her significant contribution in organising and preparing just about every aspect other of the conference.

The first Australasian Conference for Mathematics and Computers in Sport was held in 1992, and we enjoy our 20th birthday. We hope that everyone will enjoy themselves and will find the conference interactions rewarding and inspiring.

Assoc. Prof. Anthony Bedford

Assoc. Prof. Denny Meyer

MATH Sport

Scientific Committee

The organisers sincerely thank the following members of the scientific committee for their ongoing support of the conference, and reviews of papers.

Emeritus Professor Neville de Mestre
Professor Stephen R Clarke, Swinburne University
Professor Ray Stefani, California State University
Associate Professor Anthony Bedford
Associate Professor Steven Stern
Associate Professor Denny Meyer
Associate Professor Tim Heazlewood
Dr Adrian Schembri
Dr Tristan Barnett

TABLE OF CONTENTS

THE ROLE OF RESEARCH AND ANALYSIS IN THE AFL LAWS OF THE GAME PROCESS

Joel Bowden[a], Shane McCurry[a], and Patrick Clifton[a,b]

[a] *Australian Football League*

[b] *Corresponding author: patrick.m.clifton@gmail.com*

Abstract

The game of Australian Football is in great shape at the elite and community level, with record match attendances, TV audiences and participation levels in recent years. Some of the positive on-field trends have been influenced by continued monitoring and where necessary interventions in the form of rule and interpretation changes to enhance the appeal of the game. The charter of the Laws of the Game Committee is to keep the game entertaining and exciting, and safe to play within the confines of a body contact sport. Over the last 10 years, rule changes have been introduced to make the game more continuous in line with supporter expectations, to protect players from injury, and to enhance some of the traditional aspects of the game such as marking contests. The Laws of the Game process has recently undergone several positive changes; including an expansion of the personnel associated with the Laws Committee, and the introduction of a rigorous consultation and stakeholder engagement strategy. An integral part of the laws process is the extensive research and analysis that is fed into the process to assist the committee in their deliberations on various topics. This includes the publically released annual GPS and Injury Reports, Champion Data statistical reports, plus analyses of game speed, structure and fairness.

WHO REALLY WON THE 2012 MASTERS? RANKING PLAYER PERFORMANCE AND HOLE DIFFICULTY IN A MEDAL PLAY GOLF TOURNAMENT

Steven E. Stern [a,b]

[a] *The Australian National University, Canberra, Australia 0200*
[b] *Corresponding author: Steven.Stern@anu.edu.au*

Abstract

As with most individual sports, official golf rankings are based on a player's tournament placings with relative weightings for these outcomes based on the perceived importance of individual tournaments and as well as how recently the tournament was played. However, details of the performances, in terms of winning margins and aggregate scores, are not taken directly into account. Herein, we present a semi-parametric Bradley-Terry type strength estimation model for assessing the individual performances of golfers within a medal play tournament. The structure of the model takes into account actual scores on individual holes and can incorporate potentially important covariate factors such as tee-times. Further, the model structure also allows for measurement of individual hole difficulties, and thus allows an assessment of which players were potentially favoured by conditions. In so doing, we can rank player performance within a medal play tournament and assess whether the player who performed "best" actually won the tournament. The technique is applied to the tournament scores of all players in the 2012 Masters, which was won by Bubba Watson in a play-off, to assess the correspondence between actual placing and underlying player strength. In so doing, we also make comparison of the differences in strengths for players in various finishing positions with respect to the schedule of points awarded in the official rankings procedure. Ultimately, such strength estimates for performance within individual tournaments could be combined across the tournament schedule and used to derive more accurate and appropriate rankings.

Keywords: Bradley-Terry model, exponential tilting, semi-parametric, strength estimation

1. INTRODUCTION

Official rankings, both for individuals and teams are constructed across the full spectrum of sporting activities. However, the actual purposes, and therefore structure and methods, of these rankings can vary markedly (Stefani, 1999, 2010). In particular, the use of rankings, especially in individual sports, as a mechanism to encourage the most marketable players to compete widely is an often inherent aspect of many ranking systems. While it is understandable that sporting leagues have an interest in encouraging players to participate widely, this is often at odds with a "true" measure of the quality of performance, and thus with a "strength-based" ranking methodology. In this paper, we focus on golf, where the official rankings are determined according to a points accumulation system based on tournament placings over a moving two-year window (Official World Golf Rankings, 2012)

While this approach does focus on performance to an extent, it generally ignores issues such as "margins of victory", "strength of field" and "difficulty of course". Herein, we employ a newly created technique, named "asymmetric paired comparisons for ordinal outcomes", to assess the actual level of performance of golfers within a single medal play tournament. The approach, outlined in the following sections, is an extension of the standard Bradley-Terry (1952) model structure, and is applied directly to the hole-by-hole outcomes of all players in a tournament. We apply the technique to the outcomes of the 2012 Masters, and investigate the relationship between measured strength of performance and actual tournament outcomes. In this way, we should be able to eventually arrive at a ranking based on true strength of performances.

2. METHODS

The estimation of individual player strength is accomplished through the use of a new method named "asymmetric paired comparisons for ordinal outcomes", which is based on the general concept of the Bradley-Terry paired comparison model structure. Bradley and Terry (1952) proposed a model structure to assess the individual strengths, θ_i, of a collection of $i = 1,\ldots,K$ competitors which provided response data in the form of outcomes of a series of paired comparisons in which one competitor won. Specifically, they modelled

$$\Pr(\textit{Competitor i defeats Competitor j}) = H(\theta_i - \theta_j)$$

for some appropriate function $H(x)$, most commonly either a logistic function or a probit function.

In the case of medal play golf tournaments, competitors do not directly compete against one another, but instead compete directly against the golf course, the winner of the tournament then being decided by who achieved the best performance as measured against the par score. Given this extra complication, we need to modify the basic structure of the Bradley-Terry model to accommodate the structure of a medal play golf tournament as follows:

1. Each player is assigned an individual strength parameter, θ_i. [To maintain model identifiability, the sum of the player strengths is set equal to 0, meaning an "average" player has $\theta_i = 0$.]
2. Holes are assigned difficulty parameters, η_{jk}, where j represents the round number (j=1,...4), k the hole number (k=1,...,18). [Again, the η_{jk} are assumed to sum to zero for identifiability.]
3. The distribution of outcome (in terms of the score against par) for any combination of player, i, and hole, jk, is determined via an exponential tilting of a base distribution, $p_{base}(x)$, as:

$$p(x;\theta_i,\eta_{jk}) = p_{base}(x)e^{(\theta_i-\eta_{jk})x} / m(\theta_i - \eta_{jk}) \quad [1]$$

where $m(t)$ is the moment generating function (mgf) of the base distribution, which is in turn determined by the aggregate performance of all golfers on all holes in all rounds.

4. The estimated values for player strengths and hole difficulties are determined via maximum likelihood estimation; that is, they are chosen to maximize

$$L(\theta,\eta) = \prod_{i=1}^{I}\prod_{j=1}^{4}\prod_{k=1}^{18} p(x_{ijk};\theta_i,\eta_{jk})$$

where x_{ijk} is the score of player i in round j on hole k.

2.1 The base distribution

Initially, a number of standard distributions were examined as simple choices for the base distribution, including binomial, beta-binomial, hypergeometric and negative binomial structures, as well as some "zero-inflated" versions of these distributions. However, none of these choices were able to adequately capture the key structural aspect of outcomes to par on individual holes; namely, the extreme under-dispersion around par. Nearly 60% of all outcomes on any hole were pars. By contrast, those scores which were not pars had a relatively wide dispersal from double eagles to quadruple bogeys. Of course, this level of dispersal must be interpreted relative to the fact that the Masters is a major golf tournament and is thus likely to be set up to induce more dispersion of individual scores. Nevertheless, the rather unique shape of the distribution of individual hole scores needs to be properly accounted for. Therefore, it was decided that the most appropriate structure for the analysis model to take was of the semi-parametric form given above, whereby the $p_{base}(x)$ probability mass function was simply taken to be the empirical distribution of scores to par across all golfers and holes.

In the analysis of results, we undertook to examine not just overall outcomes of the entire tournament, but also the relative correctness of the mid-tournament cut as well. As such, we needed to use two different base distributions. For the overall analysis, the base distribution was constructed from all outcomes in all four rounds of the tournament and is shown in the first column of Table 1. In addition, the base distribution from only on the first two rounds, for use in investigating pre-cut performances, is provided in the second column of Table 1.

Outcome	Base Distribution Probability	
	All Rounds	Pre-Cut Rounds
Double Eagle (−3)	0.02%	-
Eagle (−2)	0.49%	0.38%
Birdie (−1)	16.84%	16.13%
Par	59.55%	59.40%
Bogey (+1)	20.36%	21.15%
Double Bogey (+2)	2.39%	2.55%
Triple Bogey (+3)	0.32%	0.35%
Quadruple Bogey (+4)	0.04%	0.03%

Table 1: Base distribution of outcomes

2.2 Tee-times and other effects

In most medal play tournaments, tee-times in the first two rounds, prior to the cut, are organised so that those players who play early on one day are scheduled to play later on the second. The primary reason behind this approach is to adjust for any differential effect in the level of difficulty of various holes at various times of day (e.g., holes in the afternoon have tended to dry out, making the greens harder and thus making approach shots more difficult to judge). Table 2 displays the average score on each hole during the first two rounds at the 2012 Masters broken down by player starting times. Early tee times were defined as those before 10am and late tee times were those after noon.

Hole	Players' Tee Times			Hole	Players' Tee Times		
	Early	Mid	Late		Early	Mid	Late
1	0.49	0.40	0.23	10	0.25	0.20	0.34
2	−0.37	−0.31	−0.30	11	0.38	0.43	0.29
3	0.02	−0.11	−0.17	12	0.18	0.26	0.00
4	0.25	0.18	0.17	13	−0.18	−0.22	−0.51
5	0.23	0.28	0.13	14	0.11	0.17	0.08
6	0.23	0.20	0.22	15	−0.32	−0.23	−0.32
7	0.36	0.38	0.05	16	0.15	0.03	0.05
8	−0.09	−0.06	−0.24	17	0.05	0.14	0.15
9	0.32	0.26	0.42	18	0.45	0.43	0.14
Out	1.34	1.23	0.52	In	1.06	1.22	0.22

Table 2: Average Pre-Cut Score to Par by Tee Times.

There appears to be an effect of tee time on the relative difficulty of holes. In particular, with the exception of the 9th and 10th, holes tend to be easier later in the day, and the aggregate front nine ("Out") and back nine ("In") scores confirm this trend.

So, to account for this, we can include covariate information in the model structure [1] by replacing individual hole difficulty parameters, η_{ij}, by a linear function of appropriate covariates for each player, $u_i = (u_{1i},\ldots u_{di})$. So, to account for tee times, we have:

$$\eta_{jk}(u_i) = \alpha_{jk} + \beta_{jk}u_{1i} + \gamma_{jk}u_{2i}$$

where u_{1i} is the indicator of whether player i had a late morning tee time and u_{2i} is the indicator of whether player i had an afternoon tee time.

Further covariate information, such as weather delay indicator variables, can then be readily included in the obvious analogous fashion.

2.3 Adjusted Scores

Once hole difficulties and player strengths have been estimated from the observed scores, an adjusted score can be constructed to facilitate fair comparisons between players. In particular, note that a sensible adjusted score is given by:

$$S_i = Hm'(\theta_i) \Big/ m(\theta_i)$$

where H is the number of holes (i.e., $H = 36$ for an adjusted score after the pre-cut rounds and $H = 72$ for an overall adjusted score), and $m'(t)$ is the derivative of $m(t)$. This adjusted score, S_i, is the estimated score to par for player i under the assumption that all the holes they played were of average difficulty (i.e., $\eta_{ij} = 0$ for all i,j).

In addition, we can examine the change in true difficulty of the course over the rounds and through the day by creating an adjusted par for the course as:

$$P_{j1} = 72 + \sum_{k=1}^{18} m'(-\alpha_{jk}) \Big/ m(-\alpha_{jk})$$

$$P_{j2} = 72 + \sum_{k=1}^{18} m'(-\alpha_{jk} - \beta_{jk}) \Big/ m(-\alpha_{jk} - \beta_{jk}) \quad [2]$$

$$P_{j3} = 72 + \sum_{k=1}^{18} m'(-\alpha_{jk} - \gamma_{jk}) \Big/ m(-\alpha_{jk} - \gamma_{jk})$$

where P_{jl} is the adjusted course par for round j and tee time l ($l = 1$ for early morning, $l = 2$ for late morning and $l = 3$ for afternoon times).

3. RESULTS

We now apply the methodology to the scores of the 2012 Masters tournament, which was eventually won by Bubba Watson in a sudden death playoff on the second extra hole. In the analysis that follows, we exclude the playoff hole outcomes (and we'll see that they actually weren't really necessary!). Table 3 shows the actual as well as adjusted scores for the top 10 players as well as the 9 players on either side of the cut-line at the end of the second round.

Top 10			Cut Line		
Player	Actual	Adjusted	Player	Actual	Adjusted
Dufner	–5	–3.90	Cantlay	+5	+4.42
Couples	–5	–3.87	Fernandez-Castano	+5	+4.42
Oosthuizen	–4	–4.01	E. Molinari	+5	+4.42
Westwood	–4	–2.94	Kraft	+5	+4.76
Garcia	–4	–2.90	Immelman	+5	+5.40
B. Watson	–4	–2.90	Cabrera	+5	+5.70
McIlroy	–4	–2.90	Bjorn	+5	+5.70
Kuchar	–3	–2.77	Kim	+6	+6.35
Jimenez	–3	–2.77	Senden	+6	+6.53
Lawrie	–3	–2.48			

Table 3: Selected actual and adjusted scores at the cut.

Note that Oosthuizen was actually "leading" at the mid-way point, so the fact that he was in the playoff at the end of the tournament should not come as a surprise. Also, note that the adjusted scores would have determined the same cut as the actual scores did. Jason Day had a score of +5 when he withdrew in the middle of the second round. His adjusted score, based on his estimated strength, was +8.43, and thus it seems unlikely that he would have made the cut had he completed his second round.

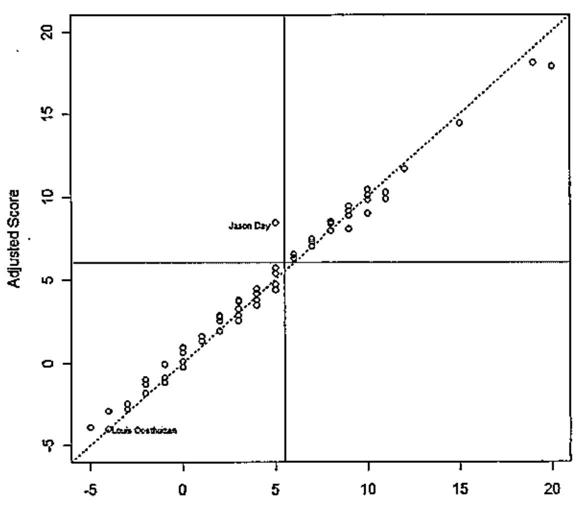


Figure 1: Pre-cut actual versus adjusted scores

For a more complete investigation, Figure 1 displays a plot of actual versus adjusted scores at the end of the second round for all 95 players. The red vertical line indicates the actual cut score, while the green horizontal line indicates the cut score that would have resulted from the adjusted scores.

As previously noted, apart from Jason Day who did not complete his second round, the players cut would have been the same based on either their actual or their adjusted scores. Also, the adjusted scores give us a way of more continuously determining placings and avoid the large number of tied positions, as they are not as discrete as the actual scores (among the 63 players who finished the tournament, there were 58 separate strength estimates but only 23 distinct finishing scores between 10 under par and 18 over par).

Player	Actual	Adjusted	Player	Actual	Adjusted
B. Watson	–10	–13.78	Bae	+4	+1.73
Oosthuizen	–10	–13.35	Johnson	+3	+1.76
Hanson	–8	–11.90	Cabrera	+3	+2.17
Mickelson	–8	–11.90	Simpson	+6	+2.61
Westwood	–8	–11.44	Bradley	+2	+2.67
Kuchar	–8	–10.59	Fowler	+2	+2.86
Poulter	–5	–8.65	McIlroy	+5	+3.16
Harrington	–4	–6.99	Baddeley	+5	+4.47
Rose	–4	–6.99	Bjorn	+4	+4.63
Furyk	–3	–6.78	Haas	+4	+4.76
Mahan	–2	–6.30	Matsuyama	+9	+4.99
Couples	–2	–6.30	Donald	+3	+5.17
Garcia	–2	–6.30	Kaymer	+6	+5.18
Scott	–4	–5.40	Chappell	+6	+5.77
Crane	–1	–4.93	Woods	+5	+5.85
Howell	E	–4.01	Jimenez	+10	+6.03
Jacobson	E	–4.01	Fisher	+7	+6.12
Lawrie	+1	–3.69	Cantlay	+7	+6.33
F. Molinari	E	–3.11	Toms	+8	+7.17
Dufner	+1	–3.10	Stricker	+7	+7.61
Snedeker	E	–2.90	Karlsson	+8	+8.63
Byrd	+2	–2.18	Schwartzel	+8	+8.69
Ogilvy	E	–1.98	Verplank	+9	+10.16
McDowell	–2	–1.76	Cink	+8	+10.17
Na	–2	–1.76	Yang	+11	+10.20
Stallings	+2	–1.50	Laird	+11	+10.44
O'Hair	+3	–1.27	E. Molinari	+11	+11.59
Watney	+3	–0.34	Woodland	+12	+14.11
Van Pelt	–1	+0.13	Fernandez–Castano	+14	+14.36
Hansen	+1	+0.47	Immelman	+13	+14.43
Singh	+2	+0.82	Kraft	+18	+19.24
Stenson	+5	+1.49			

Table 4: Actual and adjusted overall scores.

4. DISCUSSION

Table 4 provides a complete list of the actual and adjusted scores at the end of the tournament (excluding the playoff holes) for all 63 players that made the 36-hole cut (ordered according to their estimated strengths. While most adjusted scores are very similar to the actual scores, there are some notable exceptions. For example, Henrik Stenson played better than his actual score implies, which seems to be driven by his propensity for occasional disaster holes. Indeed, in his 72 holes he made 14 birdies and 2 eagles, but also a triple and quadruple bogey. There were only 2 quadruple bogeys in the whole tournament, the other by the amateur Patrick Cantlay. Interestingly, Cantley was low amateur with an overall score of 7 over par, despite Hideki Matsuyama having a better strength estimate. This result seems to be driven by the fact that Cantley had two eagles, whereas Matsuyama had none, but Matsuyama was clearly the more consistent player as he had 46 pars and only 16 holes over par compared to Cantlay's 21 holes over par. Another pair of examples of disparity between actual and adjusted scores are Adam Scott and Bo Van Pelt, who scored better than their strength would indicate mostly on the back of their aces at the 16th hole, the only two holes-in-one in the entire tournament.

In addition to the overall tabulations, we can also examine individual player performances on a hole-by-hole basis by plotting their actual scores to par on each hole played relative to the hole difficulty. As an example, Figure 2 examine Miguel Jimenez

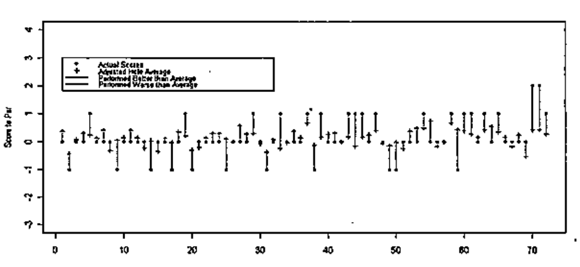


Figure 2: Hole-by-Hole Performance for Miguel Jimenez

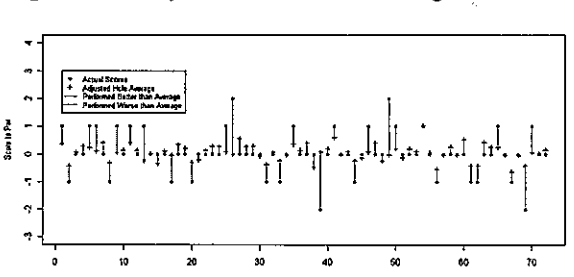


Figure 3: Hole-by-Hole Performance for Luke Donald

Note that Jimenez made a number of bogeys and two double bogeys, but these tended to be on the harder holes. Thus, his strength estimate (and associated adjusted score) is higher than would be naively guessed on the basis of his overall score of 10 over par. By contrast, Figure 3, which shows the corresponding plot for Luke Donald, indicates that, while he made fewer bogeys than Jimenez, he tended to make them on easier holes, and this is why his strength score is lower than might have been anticipated given his overall score of 3 over par.

Finally, note that these plots use adjusted average hole scores which are based on the difficulty estimates for each hole, round and tee time actually faced by the player. The range of these difficulty levels provides a means of assessing which players had the advantage or disadvantage of playing the course at harder or easier scoring periods. For the sake of clarity, we can further summarise this information by compiling the "effective" par score for each round and tee time grouping as outlined in formula [2] given at the end of Section 2. Table 5 presents these adjusted course pars along with the actual average scores of the players with the corresponding tee times.

	Tee Times		
	Early Morning	Late Morning	Afternoon
Round 1:			
Actual	74.46	73.63	71.90
Adjusted	73.40	73.25	73.26
Round 2:			
Actual	74.34	75.26	73.60
Adjusted	74.52	75.00	73.43
Round 3:			
Actual	74.67	73.05	72.38
Adjusted	74.05	73.65	74.98
Round 4:			
Actual	72.20	72.46	73.40
Adjusted	71.40	73.07	76.56

Table 5: Adjusted course pars and actual average scores by round and tee time.

As noted in Section 2.2, the actual average scores by round and tee time indicate a slight advantage to playing in the afternoon for the first two rounds. However, the adjusted scores no longer bear this out for the first round. Moreover, of the 31 players who did not make the cut, only 6 (19.4%) played in the afternoon of Round 2, while 10 (32.3%) played in the early morning and 15 (48.4%) in the mid-morning. While not statistically significant, this deviation from uniformity is indicative of how the actual scores give the appearance of an easier course in the afternoon which was not actually the case. Further, in the third round the afternoon session again seemed to be easier. Again, though, the adjusted scores do not support this perception, and this latter impression is further bolstered by the fact that the tee times in the final two rounds are determined by placings, meaning that the stronger golfers (as determined by their play in this tournament) are teeing off in the afternoon session.

Finally, based on adjusted averages, the course got harder on the last afternoon; however, better players were on course, which is why the actual averages do not present the same pattern as the adjusted scores. So, it seems early morning players may have an advantage in the final round. Is playing in later groups after the cut a disadvantage, then? If so, it pays not to be the third-round leader! Of course, at least some of the afternoon difficulty associated with the final round is associated with the pressure of winning a major championship like the Masters.

5. CONCULSION

Overall, the strength estimates from the asymmetric paired comparisons model bear out both the tournament placings and the mid-tournament cut. Moreover, they give better insight into actual player performance over the tournament, and provide a much less discrete ordering structure (indeed, using them, the victory could have been awarded to Bubba Watson without the need for a playoff!).

Further, the strength measurements could readily be combined across tournaments, via weighted averages, to arrive at a more appropriate ranking mechanism than the currently used official method. In particular, the weighing structure of the average strength calculation could incorporate the down-weighting of tournaments into the past as well as the assignment of relative importance to tournaments. The benefits to such an approach are readily seen in comparison to the points awarded for placings in the current ranking methodology. For the Masters, the winner receives 100 points while the second, third and fourth place finishers receive 60, 40 and 30 points respectively. While the large difference in points for the winner may be arguably justifiable, the fact that the fourth place getter receives only half as many points of the second place getter (and the ninth place getter receives half as many again) seems to heavily overweight the top end. By comparison, the differences in strength estimates for the players in the top ten positions are shown Table 6. The drop in strengths for the top ten places is more linear in structure than the associated drop in points awarded in the official rankings methodology, suggesting that the official approach gives too much weight to the top few placings.

Player	Position	θ_i	Player	Position	θ_i
B Watson	1	−0.567	Kuchar	T3	−0.472
Oosthuizen	2	−0.554	Poulter	7	−0.415
Hanson	T3	−0.511	Harrington	T8	−0.366
Mickelson	T3	−0.511	Rose	T8	−0.366
Westwood	T3	−0.497	Scott	T8	−0.320

Table 6: Top ten place getter strengths in 2012 Masters.

Finally, some care should be taken in combining strength estimates calculated within individual tournaments. The primary issue is the scale of the strength estimates will not necessarily be consistent across tournaments due to differences in the composition of entrants. One approach to dealing with this issue would be to somehow "standardise" the strength estimates before combining them. This could be accomplished by rescaling the strengths so that the strength of the winner is always set to a pre-specified constant value. Such an approach would also standardise the contribution of winning, which may be desirable. Alternatively, the model structure could be extended to include multiple tournaments. Doing so would be relatively straightforward from a theory perspective, however, the computational requirements of solving for the estimated strengths in such a model could well be prohibitive.

Acknowledgements

Thanks goes to Professor F. Doug Foster of the Business School at the University of Technology, Sydney for enthusiastic and useful discussions.

References

Bradley, R.A. and Terry, M.E. (1952). Rank analysis of incomplete block designs, I. the method of paired comparisons. *Biometrika, 39,* 324–345.

Official World Golf Rankings (2012). *How the System Works.* <http://www.officialworldgolfranking.com/about_us/default.sps?iType=425>

Stefani, RT. (1999). A taxonomy of sports rating systems. *IEEE Trans. On Systems, Man and Cybernetics, Part A, 29,* 116-120.

Stefani, RT. (2010). A world of sports and rating systems. *Proceedings of the10th Australasian Conference on Mathematics and Computers in Sport.* MathSport (ANZIAM), pp. 1-12.

WHO'S DRIVING THE TANK? A DYNAMIC NON-EQUILIBRIUM SIMULATION MODEL TO TEST ALTERNATIVE AMATEUR DRAFT SYSTEMS FOR MAJOR SPORTING LEAGUES

G.N. Tuck [a], A.R. Whitten [a, b]

[a] *CSIRO Mathematics Informatics and Statistics, GPO Box 1538 Hobart, Tasmania 7000, Australia*

[b] *Mezo Research, Carlton North, Victoria 3054, Australia*
[a] *Corresponding author: geoff.tuck@csiro.au*

Abstract

Teams in major sporting leagues primarily obtain amateur players through an annual draft. One of the principal drafting systems in use is the reverse-order draft, whereby poor performing teams have an opportunity to obtain the perceived best young players available. The apparent ability to improve competitive balance has led several professional sporting leagues, including the AFL, NBA, NFL and MLB, to adopt reverse-order drafts, or variants. Unfortunately, reverse-order drafts also create incentives for teams to deliberately under-perform (tank) due to the potential long-term gain from obtaining quality players at higher draft picks. Tanking – whether real or imagined – is a major concern for the integrity of sporting leagues and many leagues, such as the NBA, have a long history of adopting strategies to reduce tanking. However, methods to quantify the ability of alternative draft systems to reduce incentives to tank have not been available to sporting managers. This paper describes a dynamic, non-equilibrium simulation model that captures key components of a sporting league, including the amateur draft, player and team productivity, draft player choice uncertainty, and between-team competition. The model is used to explore how competitive balance and incentives to under-perform vary under alternative draft systems, such as reverse-order drafts, weighted drafts and draft lotteries. Results conclude that reverse-order drafts create the largest incentive to tank of all draft systems considered, but that this incentive is influenced by the ability of clubs to forecast quality players in the draft. The paper illustrates how models of this kind can be used to test alternative draft systems and shows that combinations of alternative draft systems may provide an acceptable compromise between reducing incentives to tank and improved measures of competitive balance.

Keywords: Amateur draft, competitive balance, tanking, player productivity, simulation model

Introduction

Tanking is the controversial tactic of deliberate under-performance employed by sporting clubs to gain an advantage through the drafting system of their league. This advantage is created through the provision of higher draft picks (better players) for poor performing clubs in leagues that use reverse-order draft systems, or their variants (Taylor and Trogdon, 2002; Bedford and Schembri, 2006; Kahane, 2006; Price et al., 2010). Although tanking is generally denied by sporting clubs, the topic is often raised by the media and the public as a season nears completion, and as clubs that will not make the finals look to the future.

Reverse-order draft systems are common across major sporting leagues, including the National Football League (NFL), Major League Baseball (MLB), National Hockey League (NHL), National Basketball League (NBA) and Australian Football League (AFL). Reverse-order drafts were originally introduced to improve competitive balance across competing clubs; specifically, to ensure wealthy clubs did not dominate both on and off the field and to ensure games were as competitive as possible (i.e. to ensure team strength remains reasonably even across a league) (Grier and Tollison, 1994; Berri et al., 2011). From an economic perspective, reverse-order drafts also place the power back into the hands of the club, as opposed to the player (who would otherwise simply negotiate the largest salary from the wealthiest club possible) (Dietl et al. 2011; Berri et al. 2011).

Unfortunately, reverse-order draft systems create an incentive for teams to deliberately under-perform toward the end of a season. Poor performing teams are rewarded with access to the best players available in the subsequent draft. Several studies have considered the evidence for tanking in major sporting leagues (Taylor and Trogdon, 2002; Balsdon et al. 2007; Borland et al. 2009; Price et al. 2010; Walters and Williams, 2012). Taylor and Trogdon (2002) found evidence for tanking did exist in the NBA when a reverse order-draft system was in place prior to 1985. Following concerns about tanking, a lottery draft for non-playoff teams was instituted in 1985. In the lottery draft, all teams in the lottery had an equal chance of 'winning' the number 1 draft pick, and the number of picks in the lottery was equal to the number of non-playoff teams. After the final lottery draft pick, the draft returned to a reverse-order draft. In 1987, the NBA reduced the number of picks in the lottery to three. According to Taylor and Trodon (2002), evidence for tanking was no longer apparent once the draft lottery was instituted. The NBA changed the draft system to a weighted draft in 1990. A weighted draft assumes that the probability of obtaining the best draft picks is in proportion to the finishing position of each of the non-playoff teams, i.e. the lower the standing of a club, the higher the probability of obtaining the number 1 draft pick. This system is currently in use in the NBA.

In the AFL, a reverse-order draft system has been in place since 1986. While speculation of tanking is common in the AFL, Borland et al. (2009) did not find evidence for tanking in an analysis using data from 1968 to 2005. These authors speculated that the increased propensity to tank in the NBA may be due to the greater ability to determine marquee players and the smaller team size.

The examples provided for the NBA and AFL highlight that, while managers may be concerned about tanking (as it reflects poorly on the league), there is no method *ex ante* to determine the degree of incentive to tank created by different draft systems. Logically, reverse-order draft systems will provide a greater incentive to tank than lottery systems, and weighted systems may provide mid-level incentives to tank. The relative level of incentive is however unknown – and the impact of alternative draft systems on competitive balance has not yet been investigated.

Managers of sporting leagues may face several challenging questions when designing or reviewing policies and systems intended to promote competitive balance. What if an alternative draft system was proposed? How might teams respond to take advantage of the new draft system? Do the draft systems work well for different leagues with different numbers of teams and with differing abilities to pick amateur talent? Ideally, managers of major sporting leagues would have a tool that enables them to compare draft systems and quantify the incentives to under-perform, either in an absolute or relative sense. In addition, the consequent competitive balance of the league under each draft system should be considered, as the league may like to ensure fixtures are competitive and teams do not dominate championships for an extended period of time (see alternative points of view in Forrest and Simmons, 2002).

In this paper we describe a stochastic dynamic simulation model of a major sporting league (Appendix). The focus of this paper is on testing alternative draft systems, comparing incentives to tank, and relating measures of competitive balance across these draft systems.

The draft systems considered are:

1) LD – A random draft lottery for all draft picks
2) LD1 – A random draft lottery for the lower γ=8 teams only, followed by a RO draft
3) RO – A reverse order draft
4) WD – A weighted draft where, for the bottom ω teams, the probability that a team is allocated a draft pick is in proportion to their finishing position. The number of draft picks in the lottery draft is ω_n. RO draft after pick number ω_n.
5) WTD – A waiting time draft, where the teams are allocated draft picks according to the time since reaching finishing position ρ or better.

The model parameterisation is approximated to that of the AFL (prior to the inclusion of two new teams in 2011 and 2012). Specifically, we configured the model to simulate 16 teams, 40 players on a team list, a team size of 18, and a draft pool of 80 players each year. We stipulate that 5 players must be delisted each year and 5 drafted each year. It is assumed that the first draft pick is a better player (larger value of initial productivity) than the last draft pick (a linear decline is assumed; see O'Shaughnessy (2010) for an alternative relationship between draft pick and value). However, stochastic variation in player productivity is allowed through choice uncertainty (O'Shaughnessy, 2010). Namely, it is assumed that clubs are reasonably proficient at discerning good players, i.e. high draft picks are usually good players (but not always), while low draft picks are usually poor players (but not always). For a particular year, a low draft pick player may have a higher productivity than a high draft pick player if they are closer to their peak in career performance than the high draft pick player. The full model formulation is described in the Appendix.

The incentive to tank is assumed to be measured by the proportion of premierships won by a single tanking team within a league. The greater the proportion of premierships won by a tanking team, the greater the incentive to tank. Of course, in an equilibrium sense, if an incentive to tank exists for a single team, then all teams would adopt this strategy, leading to an equal share of premierships over time. This does not imply that teams will not tank, as if one team chooses not to tank then it will be at a disadvantage in comparison to all other competitors (see the Prisoner's Dilemma literature – which is outside of the scope of this paper). Pragmatically, major sporting leagues are concerned about tanking (see the NBA example of Taylor and Trogdon, 2002) and wish to reduce the temptation for teams to tank (through the draft system, or monetary and performance penalties). The rare, stochastic and opportunistic nature of tanking, and the irrational and disparate behaviour of clubs (some are more righteous than others) means that concerns regarding tanking will inevitably exist; but ideally should be minimised. The draft system employed by a major sporting league can influence tanking temptation, and we are interested in how large this temptation might be, and how it compares relative to other drafting systems.

Methods

The dynamic model

The model is described in detail in the Appendix. The basic model structure follows the dynamics below and is based upon an AFL-like league.

1) Choose an initial set of 40 players for each team at random, with each player having a draft number (between 1 and 80) and 'age' (seasons since debut; between 1 and 15).
2) A player's productivity (ability) is a function of their draft number and age, and a random variable (with greater uncertainty regarding production as draft number increases).
3) Team productivity is assumed to be the sum over the team size of the best players in the team list (El Hodiri and Quirk, 1971; Borland, 2005; Berri et al., 2006).
4) A team with a tanking strategy reduces its team productivity by the fraction δ. However, its players do not reduce their productivity.
5) Team productivity of all teams is then compared and ranked. The team with the largest team productivity is the premier (Berri et al., 2006).
6) Players add a further year to their age, with a consequent change in their productivity.
7) Players are removed from each club either through retirement (years played greater than 15) or delisting (they have the lowest productivity on the team list).
8) Teams enter the specified draft and gain new players, with their productivity a function of draft pick and choice uncertainty.
9) Return to step (3)

The number of simulated years is 140, with summary statistics accumulated after year 40 (to remove the influence of initial conditions). A total of 3000 simulations were run for each model scenario. Within-season dynamics are not modelled; rather the finishing position of a team is assumed to be directly related to its production (or strength). Future models could consider a Tolluck-type contest function; however mean behaviour is unlikely to change within the framework adopted here. If a team adopts a tanking strategy, it does so only if in the bottom 4 teams (in terms of team productivity and therefore finishing position). Its team productivity is then reduced by δ=10% and the teams are re-ranked to decide the final finishing order of all teams. The choice of δ=10% is somewhat subjective, and analyses of the sensitivity of results to this choice could be made. The value of δ=10% was chosen as a trade-off between too small a value, whereby little gain would be obtained, and too large, whereby the clandestine nature of tanking may no longer be so. Note that the particular tanking strategy adopted by a club will be contingent on the form of the draft system and the perceived advantage gained by tanking.

The production of an individual player is assumed to be dome-shaped, with initial low production at an early stage in the player's career, through to a peak in performance, before a gradual decline until retirement (if they have not been de-listed already). Productivity functions for various sports have been estimated, and generally have a dome-shape (Schulz et al. 1994; Berri and Schmidt, 2010; Hakes and Turner, 2011). The productivity function can be estimated using performance statistics commonly available, such as on-base percentage (baseball), points scored, assists, turn-overs (basketball), possessions, goals, tackles (AFL) or other fantasy or dream team scoring systems (Berri et al., 2006; Borland et al., 2011). We have chosen the probability density function from the lognormal distribution for the productivity function because it has the necessary dome-shape and allows asymmetry in productivity about the age at peak performance (Figure 1; Appendix). In this paper we do not attempt to fit a productivity curve for the AFL, but rather keep the parameterisation general, so the effect of differing levels of choice uncertainty can be considered (none, medium, high). The peak in performance is assumed to be approximately 6 years after debut, with the potential for a 15 year career (if not de-listed) (Figure 1).

Performance measures

Two performance measures are considered. For a particular parameter set and draft system the performance measures are:

1) the percentage of premierships won, W_{won}
2) the league evenness, as measured by the mean over all years and simulations of the coefficient of variation of the team productivities across all teams for a particular year, divided by the same metric from a random lottery draft with no tanking, $CV^* = CV / CV_{lottery}$

The first performance measure gives an indication of the likely number or proportion of premierships a team may win given a draft system and tanking strategy. The second measure is similar to, but not the same as, the Noll-Scully measure of competitive balance (Noll, 1988; Scully, 1989). As the disparity between the strength of clubs increases, the larger the mean CV will become. This is then compared with a random lottery draft (with no tanking), which in essence, acts as a reference point or control by which other scenarios can be compared.

Draft systems

The lottery draft systems (LD and LD1) allocate draft picks at the conclusion of a season at random, either throughout the draft pool (LD) or only for the first γ=8 picks, i.e. the non-playoff teams (LD1). Each team in a draft lottery (teams equal to or below γ) has an equal chance of obtaining the number 1 draft pick. If γ=16 then all teams are in the first round draft lottery, and only the lower eight teams are in the draft lottery if γ=8. For LD1, after pick γ, the draft returns to a reverse-order draft (RO).

The reverse-order draft (RO) gives draft picks to clubs in reverse-order to their finishing position, for all players in the draft pool. This is the current draft system adopted by the AFL.

The weighted draft (WD) assigns a probability of obtaining a draft pick in proportion to each teams finishing position. For example, if only the lower ω teams are included in the weighted draft, then the probability of obtaining the number one draft pick is, for teams j=1,...,ω where team ω is the bottom ranked team,

$$P(\text{Number 1 draft pick for team j}) = j / \sum_{i=1}^{\omega} i$$

The team that obtains this draft pick is removed from contention for the number 2 draft pick and each team is re-allocated a probability according to their finishing position in a similar manner (with ω reduced by one). Note that the number of draft picks available in the weighted draft does not have to equal the number of teams that can obtain them. For example, while ω teams may be in the draw, only ω_n picks may be available. In this paper, $\omega=\omega_n$ for all examples considered. After pick ω_n the draw returns to a reverse-order draft.

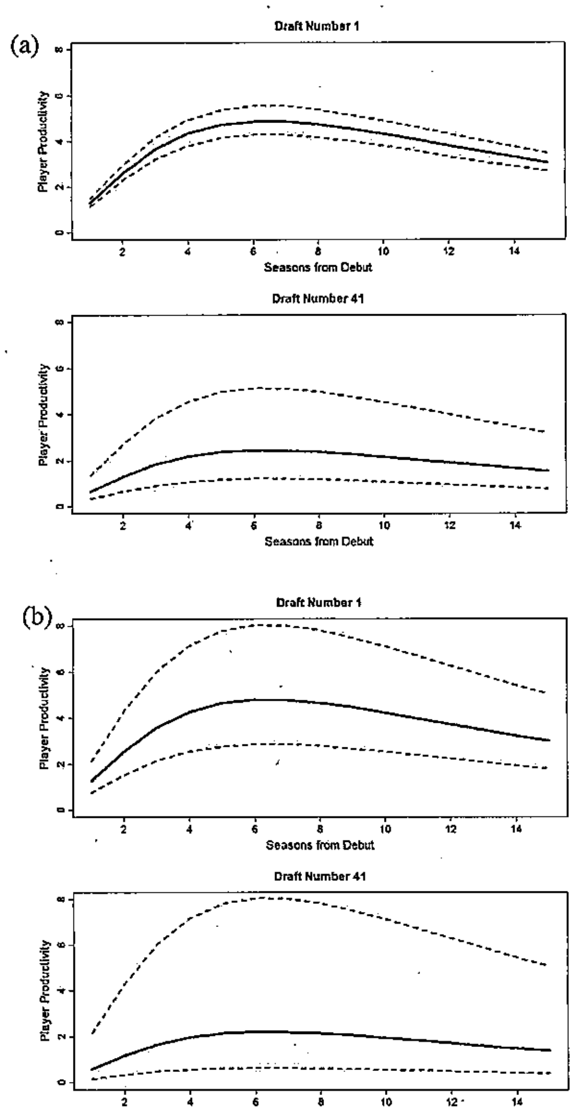


Figure 1. Player productivity as a function of seasons since debut for draft picks number 1 and 41, for (a) medium draft choice uncertainty $(\sigma_{\min}, \sigma_{\max}) = (0.05, 0.5)$ and (b) large draft choice uncertainty $(\sigma_{\min}, \sigma_{\max}) = (0.2, 0.8)$. The median player productivity is shown in black, with 99% confidence intervals in red.

The waiting time draft (WT) allocates draft picks in order of the number of years since a team has been in finishing position ρ or better. The team with the longest waiting time is allocated the number 1 draft pick, and so on. If teams have the same waiting time (for example, the top ρ teams of each season will have a waiting time of zero), they are allocated draft picks in reverse order to their finishing position. To create a history of waiting times for each club, the simulations were run for 20 years under a reverse-order draft prior to the beginning of the waiting time draft. Summary statistics were taken from year 40 onwards.

The base case parameter set is given in Table 1 of the Appendix.

Results

The expected proportion of premierships if all teams win an equal number is 1 in 16 or 6.25%. For all draft systems considered, if there is no tanking, then this is the proportion of premierships won. However, the disparity between good and poor teams does differ according to the draft system that has been invoked (Figure 2). Compared to the full lottery draft, which is being used as a control, all other draft systems showed a greater degree of evenness between teams ($CV^* < 1$). The draft with the lowest value of CV^*, and therefore the system with greatest competitive balance, was the non-playoff team lottery draft system, followed by the weighted draft and the reverse-order draft, which were very similar. The waiting time draft had the worst measure of evenness (Figure 2). As the ability of clubs to identify and subsequently pick good players decreased (choice error increased), CV^* also increased, implying that the disparity between teams becomes greater (Figure 2).

Under a full draft lottery, results show that there is no incentive to tank, as the proportion of premierships won is the same across all teams (6.25%). The waiting time draft also does not create an incentive to tank. However, the reverse-order draft, the weighted draft and the non-playoff team lottery draft all increase the proportion of premierships won for a single tanking team. The reverse-order draft shows the greatest increase in premierships won for a tanking team. The disparity between teams, CV^*, also increases if a team tanks, and this is consistent across all draft systems. As draft choice uncertainty increases, the incentive to tank decreases for all draft systems considered (Figure 2).

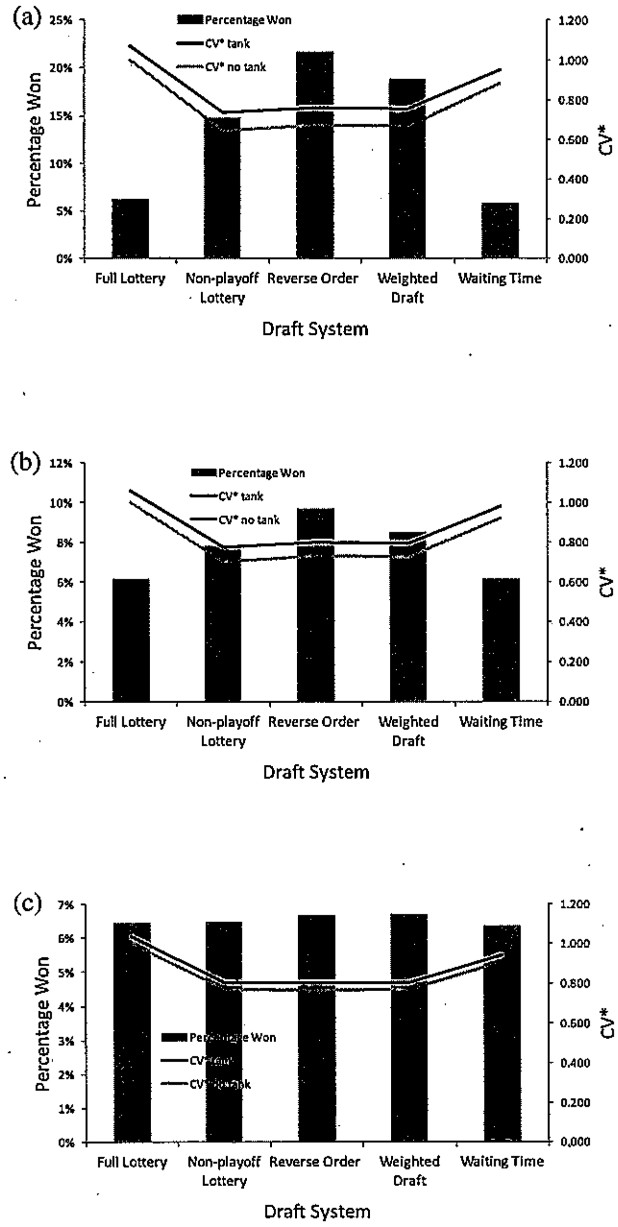


Figure 2. For each draft system, shown are the percentage of premierships won for a single tanking team and the relative coefficient of variation, CV*, for a league with a tanking team and if no team tanks. (a) No draft choice uncertainty $(\sigma_{min}, \sigma_{max}) = (0,0)$, (b) medium choice uncertainty, $(0.05,0.5)$ (c) large choice uncertainty, $(0.2,0.8)$.

For the weighted draft, as the number of draft picks included in the weighted draft ($\omega = \omega_n = 2, 4, 6, 8$) is increased, the incentive to tank decreased only marginally. With only 2 picks in the weighted draft, the percentage of premierships won for a tanking team is very similar to (but less than) a reverse-order draft (Figure 3). Under the assumed tanking structure, there is no incentive to tank created by the waiting time draft. However, as the finishing position threshold increases (ρ =2, 4, 6, 8), the disparity between teams reduces (Figure 4). For the lottery draft, the greatest incentive to tank is with the fewest (γ=2) picks in the lottery (Figure 5). This is because, other than the two picks in the lottery, it essentially resembles a reverse-order draft (Figure 2(b)). As the number of picks in the lottery increases, the incentive to tank reduces, until all players are in the lottery (γ=80), and there is no incentive to tank.

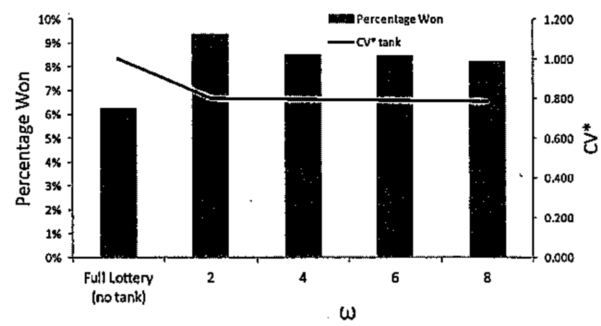


Figure 3. For the weighted draft, the percentage of premierships won and CV* for a league with a single tanking team as a function of the number of draft picks in the weighted draft, $\omega = \omega_n$, and medium draft uncertainty $(\sigma_{min}, \sigma_{max}) = (0.05,0.5)$.

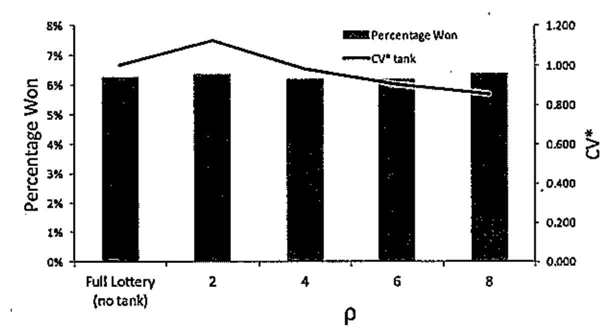


Figure 4. For the waiting time draft, the percentage of premierships won and CV* for a league with a single tanking team as a function of the finishing position threshold, ρ, and medium draft uncertainty $(\sigma_{\min}, \sigma_{\max}) = (0.05, 0.5)$.

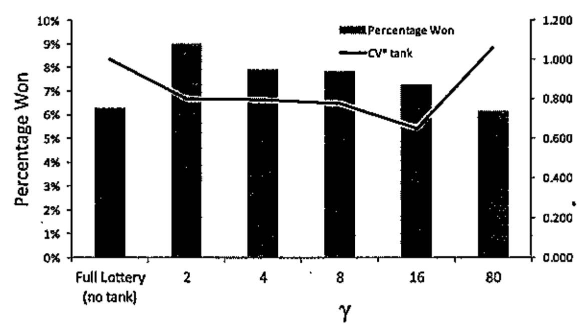


Figure 5. For the lottery draft, the percentage of premierships won and CV* for a league with a single tanking team as a function of the number of picks in the lottery, γ, and medium draft uncertainty $(\sigma_{\min}, \sigma_{\max}) = (0.05, 0.5)$.

Discussion

For resource managers, whether the resource is a financial commodity, a natural resource or a sporting team, the ability to predict the resource response prior to implementing a management strategy is of great benefit (Sainsbury et al., 2000). In sporting leagues, managers have a number of complex issues to consider regarding the long-term sustainability and health of their league and the teams. Amateur player draft systems are a commonly employed mechanism for achieving competitive balance. Unfortunately, player draft systems can be exploited by teams that deliberately under-perform to re-position them in the draft and obtain better players.

In this paper, we have shown how alternative draft systems can be designed and then tested to see how the relative incentive to tank, and the consequent evenness amongst teams, compares. In addition, key league characteristic such as the level of uncertainty in player draft choice, the number of teams in the league, the team size, *inter alia,* could be considered.

Reverse-order drafts, and their variants, are the most common player draft systems used amongst major sporting leagues. Of all the draft systems considered in this paper, the basic reverse-order draft is also shown to create the greatest incentive for teams to tank. As expected, no incentive to tank is created by a full draft lottery. However, this draft system does not enable chronically poor performing teams (whether through poor administration or by chance) any surety that they will recover. Other draft systems also have their problems as far as measures of competitive balance are concerned. For example, reverse-order drafts can lead to teams becoming stuck in the middle-ranks for extended periods of time.

The weighted draft had a lower incentive to tank than the reverse-order draft, but higher than the non-playoff team draft lottery. Weighted drafts are currently used in the NBA. However, the weighted draft in the NBA differs from that described here. Because the probability of obtaining the number 1 draft pick for higher ranked (better) teams was perceived by the NBA to be too large, in 1993 the system was changed so that teams with a poorer record have a higher probability of obtaining the top draft picks (picks 1 to 3 only). In the weighted draft here, the number of draft picks available was also equal to the number of teams that could obtain them ($\omega = \omega_n$).

The waiting time draft was created to reduce the time it takes for poor performing teams to reach a satisfactory level of on-field success. This system ensures, through the provision of high draft picks, that teams do not become stuck in mid-rank and essentially forces them into success. We are not aware of any sporting league that has adopted this draft system. One of the disadvantages of this system is that it provides the number 1 draft pick to the team with the longest waiting time until it eventually reaches or exceeds the waiting time finishing position threshold, ρ. This is unlikely to be popular amongst clubs and sporting fans as it clearly favours certain teams (in terms of top draft picks) over an extended period of time. The provision of numerous high draft picks also appears to increase the disparity between teams, CV^* (Figure 2). However, it does not create any incentive to tank, as draft picks are not necessarily given to the bottom teams. Even an alternative tanking strategy that loses games to retain the number 1 draft pick is not likely to be popular, as it would mean deliberately losing a final or playoff game (if managers have chosen ρ so that it is less than the number of teams playing in the finals).

Potential draft strategies that follow from these analyses, and might be acceptable to managers, could combine draft systems with favourable aspects. For example, a new draft could include a non-playoff team draft lottery combined with a waiting time draft. The draft lottery reduces the tanking incentive (over a reverse-order draft) and has a low CV^*, while the waiting time draft has no incentive to tank and reduces the time for poor performing teams to reach the finals. With the waiting time draft following the draft lottery, clubs may be less concerned about teams being consistently provided with high draft picks. Alternatively, a draft system could have a random or weighted draft lottery for teams (say 4 teams) with the longest waiting time, followed by a waiting time draft (or reverse-order draft). This removes the concern that a single team consistently obtains the number 1 draft pick (until it reaches ρ) with the waiting time draft. While these draft system have not been tested in this paper, it illustrates how draft systems can easily be created and their utility explored.

Several alternative drafting systems have been considered in this paper, and others could be formulated and examined (Bedford and Schembri, 2006; O'Shaughnessy, 2010). However, only one tanking strategy was tested, namely a single team tanks if in the lower 4 teams. The framework established in this paper not only allows alternative draft systems to be compared, but also alternative tanking strategies. For example, with a lottery or weighted draft where only the bottom four teams have access to the number 1 draft pick (γ=4; ω=ω_n=4) a strategy whereby a team tanks if in positions marginally above fourth from the bottom would allow that team potential access to the number 1 draft pick. Further analyses should consider the weaknesses of each of the draft systems and then impose tanking strategies on those systems. In this way, the relative strengths and weaknesses of the drafts can be tested.

Future models, where more direct applications to particular sporting leagues are necessary, should consider fitting data to the productivity functions. This paper instead considered a broad range of potential uncertainty for the productivity function (that may capture the actual uncertainty of the league in question). In this way, a more thorough exploration of the model behaviour was possible. It was found that the larger the choice uncertainty, the smaller the incentive to tank. If clubs have little confidence in their ability to pick marquee players, even with the number 1 draft pick, then the incentive to tank will diminish. However, if clubs are not able to adequately pick quality players, then the basic point of draft systems, such as the reverse-order draft, is lost. Likewise, if players are traded away from poorer clubs and bought by richer clubs soon after drafting, then reverse-order drafts, once again, are compromised (Booth, 2004; Dietl et al., 2011). In this respect, the model described here is more suited to the AFL (a league of win-maximising clubs) than many of the North American sporting leagues, as the AFL has tighter controls that influence player movement, and mass trading between clubs remains the exception rather than the rule (Booth, 2004; Vrooman, 2009). However, refinements of the model could include player trading and economic factors, such as salary caps, and penalties for tanking. In addition, terms more familiar with profit-maximising clubs, such as gate revenue and player value could also be included (El Hodiri and Quirk, 1971; Quirk and El Hodiri, 1974; Fort and Quirk, 1995). The impact of these economic factors on incentives to tank could then be considered relative to the draft system employed and the financial status of the club. However, in a qualitative sense, we suspect that many of the results illustrated in this paper are likely to hold.

Conclusion

This paper shows how alternative draft systems for major sporting leagues can be assessed in terms of the incentives they create for deliberate under-performance and consequent competitive balance.

An additional important point of this paper is that whole-of-system models involving complex dynamics can be used to assist managers when making difficult decisions that often involve trade-offs between conflicting objectives (Sainsbury et al. 2000; Millner-Gulland, 2011). It is clearly beneficial for managers to be able to predict resource responses to management strategies prior to their implementation. Modelling frameworks of this kind can be used as decision tools, with club administrators, managers and analysts directly involved in the process of developing the management strategies (such as alternative draft systems), assigning performance measures, and crafting behavioural responses that attempt to exploit weaknesses in the management. In this way all key stakeholders are engaged and have ownership of the management strategies that are (eventually) adopted. While this process is used in natural resource sciences (Sainsbury et al., 2000; Millner-Gulland, 2011), in this paper we have used a hypothetical example, based upon the Australian Football League, to illustrate how it can be utilised by major sporting leagues.

Acknowledgements

We wish to thank Natalie Dowling (CSIRO) for her helpful comments on an earlier draft of this paper, and two anonymous referees for their valuable suggestions. Thanks are due to Peter Campbell (CSIRO) for his assistance with scientific computing. Campbell Davies (CSIRO) is thanked for his ongoing support.

References

Balsdon, E., Fong, L. & Thayer, M.A. (2007). Corruption in college basketball? Evidence of tanking in post-season conference tournaments. *Journal of Sports Economics, 8*, 19-38.

Bedford, A., & Schembri, A. J. (2006). A probability based approach for the allocation of player draft selections in Australian Rules Football. *Journal of Sports Science and* Medicine, 5, 509-516.

Berri D.J. & Schmidt, M.B. (2010). Stumbling on Wins: two economists expose the pitfalls on the road to victory in professional sports. FT Press, USA.

Berri D.J., Schmidt, M.B., & Brook, S.L. (2006). The wages of wins: taking measure of the many myths in modern sport. Stanford University Press, USA.

Berri, D.J., Brook, S.L, & Fenn, A.J. (2011). From college to the pros: predicting the NBA amateur player draft. *Journal of Productivity Analysis. 35,* 25-35.

Booth, R. (2004). The Economics of Achieving Competitive Balance in the Australian Football League, 1897-2004. *Economic Papers, 23*, 325-344.

Borland, J. (2005). Production functions for sporting teams. Working paper, Department of Economics, University of Melbourne, http://www.economics.unimelb.edu.au/staff/jib/documents/Prodfunctforsportingteams.pdf

Borland, J., Chicu, M., & Macdonald, R. D. (2009). Do Teams Always Lose to Win? Performance Incentives and the Player Draft in the Australian Football League. *Journal of Sports Economics, 10*, 451-484.

Borland, J., Lee, L., & Macdonald, R.D. (2011). Escalation effects and the player draft in the AFL. Labour Economics, 18, 371-380.

Coase, R. (1960). The problem of social cost. *Journal of Law and Economics*, 3, 1-44.

Dietl, H., Fort, R. & Lang, M. (2011). International sport league comparisons. In Handbook on International Sports Management and Policy. Ed: Downward, P. London, UK. Routledge.

El Hodiri, M. & Quirk, J. (1971). An economic model of a professional sports league. *Journal of Political Economy, 79*, 1302–1319.

Forrest, D. & Simmons, R. (2002). Outcome uncertainty and attendance demand in sport: the case of English soccer. *The Statistician*. 51, 229 - 241.

Fort, R. & Quirk, J. (1995). Cross-subsidization, incentives, and outcomes in professional team sports leagues. *Journal of Economic Literature, 33*,1265–1299.

Grier, K. & Tollison, R.D. (1994). The rookie draft and competitive balance: The case of professional football. *Journal of Economic Behavior & Organization,* 25, 293-298.

Hakes, J.K. & Turner, C. (2011) Pay, productivity and aging in major league baseball. *Journal of Productivity Analysis*. 35, 61 - 74.

Kahane L. (2006). The Reverse-order-of-finish Draft in Sports. Handbook on the Economics of Sport, Andreff W., Szymanski S. (eds.), Cheltenham, Edward Elgar, p643-5.

Millner-Gulland, E.J. (2011). Integrating fisheries approaches and household utility models for improved resource management. *PNAS*, 108: 1741-1746.

Noll, R. (1988). Professional Basketball. Stanford University Studies in Industrial Economics, No. 144.

O□Shaughnessy, D. (2010). On the value of AFL player draft picks. Proceedings of the 10th Conference on Mathematics and Computers in Sport 2010.

Price, J., Soebbing, B.P., Berri, D. & Humphreys, B.R. (2010) Tournament incentives, league policy, and NBA team performance revisited. Journal of Sports Economics, 11, 117-135.

Quirk, J. & El Hodiri, M. (1974). The Economic Theory of a Professional Sports League. In Government and the Sports Business, edited by R. Noll. Washington DC. Brookings Instituation.

Rottenberg, S. (1956) The baseball player's labor market. *Journal of Political Economy*, 64, 643-761

Sainsbury, K.J., Punt, A.E. & Smith, A.D.M. (2000). Design of operational management strategies for achieving fishery ecosystem objectives. *ICES Journal of Marine Science 57*: 731-741.

Schulz, R., Musa, D., Staszewski, J., & Siegler, R.S. (1994). The relationship between age and Major League Baseball performance: implications for development. *Psychology and Aging, 9*, 274–286.

Scully, G. W. (1989) The Business of Major League Baseball. University of Chicago Press; Chicago, IL.

Taylor, B.A. & Trogdon, J.G. (2002). Losing to Win: Tournament Incentives in the National Basketball Association. *Journal of Labor Economics, 20*, 23-41.

Vrooman, J. (2009). Theory of the perfect game: competitive balance in monopoly sports leagues. *Review of Industrial Organisation*, 34, 5-44.

Walters, C. & Williams, T. (2012). To tank or not to tank? Evidence from the NBA. MIT Sloan Sports Conference March 2-3, 2012, Boston, MA, USA.

Appendix

Player productivity

The player productivity function, $u_{d,a}$, is assumed to be dome-shaped and asymmetrical, and is defined by,

$$u_{d,a} = f(a+1;\mu_u,\sigma_u)(md+1-m)\varepsilon_d. \quad (1)$$

where

$f(a+1;\mu_u,\sigma_u)$ is the probability density function of the lognormal distribution and defines the shape of the base productivity curve (when $d=1$ and $\varepsilon_d=1$) as a function of age a $(a=1,\ldots,a_{\max})$

μ_u is the shape parameter for the base productivity function

σ_u is the scale parameter for the base productivity function

m is the slope of the linear relationship relating draft number to productivity, and is

$$m = (1-d_{\min})/(1-d_N) \quad (2)$$

where d_N is the maximum draft number, and $d_{\min}$ is the minimum value in the linear relationship occurring when $d = d_N$

$\varepsilon_d \sim LN(\mu_d,\sigma_d^2)$ is a random variate from a lognormal distribution, with the mean of ε_d equal to 1. The location parameter μ_d and scale parameter σ_d are defined by

$$\mu_d = -\sigma_d^2/2 \quad (3)$$

$$\sigma_d = \sigma_m(d-1)+\sigma_{\min} \quad (4)$$

$$\sigma_m = (\sigma_{\max}-\sigma_{\min})/(d_N-1) \quad (5)$$

where σ_m is the slope of the relationship defining the increased uncertainty with draft number,

$\sigma_{\min}$ is the scale parameter for the number 1 draft pick $d=1$

$\sigma_{\max}$ is the scale parameter for the final draft pick $d=d_N$.

The dynamic model

The model assumes that there are L teams in the league, and that each team has P players on their team list (or squad). Of the P players on the team list, p players participate in a season (the team size). The initial ages for the player productivity functions for each team, l, are chosen at random between ages 1 to $a_{\max}$ (ages are rescaled without loss of generality). The initial draft number for each player within a team is chosen at random between 1 and the maximum draft number d_N, without replacement.

The team productivity, $U^{l,y}$, for team l for a particular year y is the sum over the p highest ranked player productivities of the team,

$$U^{l,y} = \sum_{i=1}^{p} u_{i,d,a}^{l,y}, \quad (6)$$

where $u_{i,d,a}^{l,y}$ is the ranked player productivity for team l, in year y, with draft number d and age a, and the player productivity for player i is greater than player $i+1$.

The finishing positions are determined from the ranking of the team productivities. The premiership winner for the year is the team with the highest ranked team productivity. The order of allocation of draft picks is then determined by the particular player drafting system adopted by the league.

In order to make room for new players from the end-of-year draft, players can be removed from team lists in two ways, (i) player retirements and (ii) de-listing. Each team must remove a minimum of r_N players from their playing list. As there are L teams, this implies there must be at least $d_N = r_N L$ players in the draft. Firstly, players are removed that are of age greater than $a_{\max}$ (player retirements). If there are greater than r_N retirements, then all players recruited beyond the ${r_N}^{th}$ player receive the equivalent player productivity to the $(d_N)^{th}$ player drafted (with

$\varepsilon_d = 1$). If there are less than r_N retirements in a team, then players with the lowest player productivity are removed (de-listed), until there are a total of at least r_N players marked for removal. These players are then removed from the playing list and the teams then enter the player draft.

New players are assigned to each team with their attributes defined by the player productivity function (equation 1) with (i) $a = 1$, (ii) their particular draft number, d and (iii) uncertainty regarding player productivity defined by ε_d. At the end of the draft process all teams have a full team list of P players once again. The model then moves to the next year, $y+1$, player productivities are updated, and the process of summing over ranked player productivities to determine the team productivity, premiership success and drafting begins again. A single simulation concludes in year $y = Y$. To account for the potential influence of the initial conditions assigned to player productivity in year $y = 1$, each simulation is repeated $s = 1$ to S times.

Under-performance

Once a team's end-of-season team productivity has been calculated (equation 6), under-performance can be modelled by reducing the under-performing team's productivity by the fraction δ,

$$U_{\delta}^{l,y} = (1-\delta)U^{l,y} \tag{7}$$

for all teams $l = 1,\ldots,L_{\delta} \leq L$ that have adopted an under-performing strategy. The parameter δ gives an indication of the degree of tanking, that is, the larger the fraction, the more heavily a team tanks. Teams that have made the end-of-season finals have no incentive to under-perform, as might teams still vying for positions in the playoffs near the end of the season. Therefore the model assumes that a team only under-performs if its ladder position is below q.

Table 1. The base-case parameter set based upon an AFL-like competition. Note that alternative values were used to explore model sensitivity, and these values are defined in the text.

Parameter	Description	Value
S	The total number of simulations	3000
Y	The total number of simulated years	140
L	The number of teams in the league	16
P	The number of players on the team list	40
p	The team size	18
d_N	The number of players in the draft	80
r_N	The minimum number of players per team replaced each year	5
$a_{\max}$	The maximum number of seasons per player (or re-scaled retirement age)	15
$d_{\min}$	The mean reduction in productivity for $d = d_N$ relative to $d = 1$	0.05
μ_u	The shape parameter for the base productivity function	ln(14)
σ_u	The scale parameter for the base productivity function	0.8
$(\sigma_{\min}, \sigma_{\max})$	The scale parameters for $(d = 1, d = d_N)$	(0.05, 0.5)
ω	The number of teams in the weighted lottery	4
ω_n	The number of picks available in the weighted lottery	4
ρ	The waiting time finishing position threshold	4
γ	The number of teams in the draft lottery	8
q	The finishing position below which a team adopts an under-performance strategy	12 (lower 4 teams)
L_δ	The number of teams adopting an under-performance strategy	1
δ	The discount to the team productivity for a team that under-performs	0.1

A HELPING HAND TOO FAR: CAN PLAYER DRAFT CONCESSIONS DESTABILISE COMPETITIVE BALANCE IN SPORTING LEAGUES?

Athol Whitten [a,b,*], Geoff Tuck [b]

[a]*Mezo Research, Melbourne, Australia,*
[b] *CSIRO Marine and Atmospheric Research,*
[*] *Corresponding author: athol.whitten@mezo.com.au*

Abstract

Amateur player draft systems, such as reverse-order or lottery drafts, are the principal management controls used by major sporting leagues to give teams access to new young players. However, draft concessions may be necessary if teams underperform for long periods, or when new teams enter a competition. These draft concessions are provided to enable low ranking or new teams access to greater numbers of the most talented young players, in the belief that this will accelerate improvements in team strength and on-field success.

Despite the prevalence of these systems, there are few methods to assess the potential for draft concessions to increase a team's long term success, and no well-defined targets by which to measure management success. This paper implements a dynamic non-equilibrium simulation model to explore how competitive balance varies when draft concessions are provided to new or under-performing teams. By capturing key components of a sporting league; player and team productivity, between team competition, amateur player draft systems, draft choice error, and the provision of draft concessions, we demonstrate that generous draft concessions may result in disproportionate increases in success for teams with draft concessions and declining competitive balance across the league. Our results lead to several simple questions: can we predict the success or otherwise of management strategies such as the provision of draft concessions? And what exactly are the management targets?

Keywords: Player draft systems, draft concessions, competitive balance, dynamic simulation

TEAM PERFORMANCE IN THE AUSTRALIAN FOOTBALL LEAGUE

Stephen R Clarke [a,b], Rodney C Clarke

[a] *Swinburne University of Technology*
[b] *Corresponding author: sclarke@swin.edu.au*

Abstract

The best team in the AFL (1990-2011) is investigated from several angles. Many fans count the number of premierships and finals appearances as the only true measure of success. On this measure West Coast is the most successful AFL club. However this is subject to large random chances. A premiership table using all matches in the AFL is used as a more robust measure of success, with Geelong topping the table on winning percentage. The large home advantage enjoyed by the non-Victorian clubs is demonstrated by splitting this table into home and away performance in non-finals matches. Only Richmond, Fitzroy and Gold Coast failed to win more than 50% of their home matches while only Geelong and Nth Melbourne won more than 50% of away matches. An algorithm based on a successful automated tipping program of the last 30 years is used to generate a rating for each team after each match. These ratings are used to evaluate the best and worst teams of each club over the 22 years of the AFL. While Geelong had the highest mean rating, the highest ever rating was achieved by the Essendon team of 2001, followed by Collingwood 2011 and Brisbane Lions 2002. Although Gold Coast struggled in its first year, eight other clubs have at some stage had teams rating worse than the lowest rating of the Gold Coast. A rating over 124 (one standard deviation over the mean) was used as a measure of excellence – only 15% of weekly ratings exceeded this level. With 203 weeks at this level, Geelong was well ahead of West Coast with 145, Brisbane Lions with 113 and Essendon 110. Fitzroy, Richmond and Gold Coast have never achieved this level. However in terms of a continuous run over this level, Geelong have an as yet unbroken run of 114 rounds starting in round 12 of 2007, Essendon started a run of 61 rounds in round 19 of 1999, and Collingwood started an unbroken run of 47 rounds in round 5 of 2010.

Keywords: computer rating, home advantage, Australian rules

1. INTRODUCTION

The Victorian Football League (VFL) was formed in 1897 with eight Victorian based clubs; Carlton, Collingwood, Essendon, Fitzroy, Geelong, Melbourne, South Melbourne and St. Kilda. In 1908 Richmond and University were introduced in to the competition, but University withdrew in 1914 due to World War 1. Further expansion followed in 1925 with Footscray, Hawthorn and North Melbourne entering the competition making a total of 12 Victorian clubs. The competition held this structure for over half a century until 1982 when South Melbourne relocated to Sydney. Brisbane and West Coast were introduced in 1987, and in 1990 the VFL changed its name to the Australian Football League (AFL). Further expansion has occurred since, with Adelaide being introduced in 1991, Fremantle in 1995 and Port Adelaide in 1997. The start of the 1997 season saw another change, the merger of Fitzroy and Brisbane. 2011 saw Gold Coast introduced and lastly, Greater Western Sydney were introduced in to the competition in 2012. Over the life of the AFL many teams have changed names, often incorporating their club mascot, and most grounds have changed naming rights. In this paper we use generic names such as Footscray and Docklands for teams and grounds. These usually refer to locations.

In the early VFL years, each club used a unique suburban home ground, but later some teams moved to different grounds to play their home matches. Nowadays, most teams share a home ground with at least one other team and often choose to play their home matches at different grounds depending on the opposition, potential crowd and sponsorship deals. Table 1 shows each club's major home ground and the years they have played there since the start of the AFL

Club	Major Home Ground	Year
Adelaide	Football Park	1991 – 2011
Brisbane	Carrara Stadium	1990 - 1992
	BrisbaneCG	1993 - 2011
Carlton	Princes Park	1990 - 2005
	Docklands	2006 - 2011
Collingwood	Victoria Park	1990 - 1992
	MCG	1993 - 2011
Essendon	Windy Hill	1990 - 1991
	MCG	1992 - 1999
	Docklands	2000 - 2011
Fitzroy	Princes Park	1990 - 1993
	Western Oval	1994 - 1996
Footscray	Western Oval	1990 - 1996
	Princes Park	1997 - 1999
	Docklands	2000 - 2011
Fremantle	Subiaco	1995 - 2011
Geelong	Kardinia Park	1990 - 2011
Gold Coast	Carrara Stadium	2011 - 2011
Hawthorn	Princes Park	1990 - 1991
	Waverley Park	1992 - 1999
	MCG	2000 - 2011
Melbourne	MCG	1990 - 2011
Nth. Melbourne	MCG	1990 - 1999
	Docklands	2000 - 2011
Port Adelaide	Football Park	1997 - 2011
Richmond	MCG	1990 - 2011
St. Kilda	Moorabbin Oval	1990 - 1992
	Waverley Park	1993 - 1999
	Docklands	2000 - 2011
Sydney	SCG	1990 - 2011
West Coast	Subiaco	1990 - 2011

Table 1. Home grounds of each AFL club.

Fans and commentators often debate which team of a particular decade or era is the best, and there are obviously many different factors that determine this. This paper investigates which is the best performed team in the 22 years of the AFL, covering the period 1990-2011 inclusive.

The major success factor usually used in determining the best team is the number of premierships won by that team. Table 2 presents the number of premierships and finals appearances of each club in the AFL. On this basis West Coast, Geelong and Brisbane share the honours.

Clearly some clubs are more 'efficient' than others. StKilda and Footscray's 11 years in the finals have resulted in three fewer flags than Brisbane's 10 years in the finals. This may be due to a home advantage bias in the finals. Currently the higher ranked team has a home state advantage in the first three weeks of the finals. This gives interstate teams a large home advantage in the finals which many Victorian teams do not enjoy.

Club	Number of premierships	Number of years club played in finals	Total number of finals played
WestCoast	3	17	40
Geelong	3	14	38
Brisbane	3	10	25
Essendon	2	14	30
NorthMelbourne	2	12	27
Hawthorn	2	12	22
Collingwood	2	11	28
Adelaide	2	11	26
Sydney	1	13	25
Carlton	1	10	23
PortAdelaide	1	7	17
St.Kilda	0	11	25
Footscray	0	11	24
Melbourne	0	9	20
Fremantle	0	3	6
Richmond	0	2	6
Fitzroy	0	0	0
GoldCoast	0	0	0

Table 2. Frequency of premierships and finals appearances in the AFL 1990-2011

However a premiership is decided on one game, which, like any final, can be decided on one kick or one lucky bounce. This paper looks at more robust measures.

2. AFL LADDER

Although every player strives for and every fan wants their team to achieve premierships, they do not tell the whole story. Currently a team must win at least three finals matches on end to win the premiership, and factors such as injury, luck and home ground advantage can result in the dominant team of the year not winning the premiership.

Team	Played	Won	Drawn	Lost	Win%	For	Against	Point %
Geelong	520	319	4	197	61.7	54354	46679	116.4
WestCoast	522	293	5	224	56.6	49345	46387	106.4
Essendon	512	280	8	224	55.5	52016	49239	105.6
NthMelbourne	509	277	3	229	54.7	51600	50289	102.6
Collingwood	510	271	6	233	53.7	50057	46541	107.6
Hawthorn	504	260	3	241	51.9	48289	46565	103.7
PortAdelaide	347	178	4	165	51.9	32315	32670	98.9
Adelaide	486	251	1	234	51.7	45646	43853	104.1
St.Kilda	507	257	8	242	51.5	47997	47291	101.5
Footscray	506	253	8	245	50.8	49533	49605	99.9
Carlton	505	243	4	258	48.5	49092	49293	99.6
Sydney	507	237	7	263	47.4	47710	48205	99.0
Brisbane	507	233	8	266	46.7	48650	49393	98.5
Melbourne	502	216	4	282	43.4	46361	50114	92.5
Fremantle	380	153	0	227	40.3	33479	37310	89.7
Richmond	488	192	5	291	39.9	43952	50703	86.7
Fitzroy	152	38	0	114	25.0	12673	17740	71.4
GoldCoast	22	3	0	19	13.6	1534	2726	56.3

Table 3: Premiership table based on all AFL matches, 1990-2011

The authors believe that home and away matches should be taken in to account when determining the most successful club in the AFL. Table 3 presents a ladder of all the matches played since the AFL was formed, including finals. As each club has played a different number of games, it is ranked in order of win percentage. Note that win percentage is defined as 100 * number of matches won / number of matches played (counting a draw as half a win). In keeping with the usual practice in AFL ladders the (point) percentage is calculated as 100 * points won / points lost.

From this table we can see Geelong out in front having won more than 60% of their matches. Their percentage is also very impressive; ten points higher than the second placed West Coast and nine points higher than Collingwood. Unsurprisingly, Fitzroy and Gold Coast languish at the bottom of the ladder with very poor win percentages.

Brisbane, who have won three premierships, have actually won fewer games than they have lost, a big part due to their time as the Brisbane Bears where they spent a long time down the bottom of the ladder. The top six teams on the ladder have won at least two premierships, while the bottom five have failed to win any. West Coast have played the most games and hence the most finals, 34 more than Richmond since 1990.

Interestingly, of the 13 teams that have been in the competition since 1990, Hawthorn has only played more finals than Melbourne and Richmond, yet has won two premierships. St. Kilda and Footscray have each won more than half their matches, yet failed in their pursuit for a flag.

3 HOME ADVANTAGE

Home advantage in team sports is well documented in the literature. Stefani and Clarke (1992), Clarke (2005) and Ryall and Bedford (2010, 2011) have evaluated its effects in Australian Rules football. Tables 4 and 5 split the performance of all teams in the AFL when playing at home and away. Because the higher ranked team is often given the benefit of a home match in the finals, or listed as the home team on the MCG, this table does not include finals. Even so, the effects of the home advantage are underestimated, as many of these matches would be played on neutral grounds, or even on the home grounds of the nominally away side. This is due to many teams sharing grounds, and the AFL's practice of moving some matches to the MCG.

Home Team	Played	Won	Drawn	Lost	Win%	For	Against	Point %
Geelong	241	172	3	66	72.0	25713	20193	127.3
WestCoast	241	168	1	72	69.9	25046	19929	125.7
Essendon	241	158	3	80	66.2	25976	23101	112.4
Adelaide	230	152	0	78	66.1	23031	18612	123.7
PortAdelaide	165	104	2	59	63.6	15999	14167	112.9
Collingwood	241	143	1	97	59.5	24389	21385	114.0
Footscray	241	139	7	95	59.1	24746	22697	109.0
NthMelbourne	241	141	2	98	58.9	25167	23455	107.3
St.Kilda	241	139	6	96	58.9	23714	21176	112.0
Brisbane	241	138	4	99	58.1	24843	22146	112.2
Hawthorn	241	140	0	101	58.1	23381	21302	109.8
Sydney	241	133	1	107	55.4	24288	22487	108.0
Fremantle	187	102	0	85	54.5	17510	16824	104.1
Carlton	241	128	1	112	53.3	24059	22932	104.9
Melbourne	241	120	3	118	50.4	23194	23833	97.3
Richmond	241	114	3	124	47.9	22837	24409	93.6
Fitzroy	76	24	0	52	31.6	6525	8670	75.3
GoldCoast	11	1	0	10	9.1	795	1357	58.6

Table 4: Performance of AFL clubs at home in the home and away rounds 1990-2011

Away Team	Played	Won	Drawn	Lost	Win%	For	Against	Point %
Geelong	241	125	1	115	52.1	24844	23246	106.9
NthMelbourne	241	121	1	119	50.4	23897	24085	99.2
Collingwood	241	112	3	126	47.1	23141	22886	101.1
Hawthorn	241	109	3	129	45.9	22826	23314	97.9
Essendon	241	107	5	129	45.4	23184	23542	98.5
St.Kilda	241	108	1	132	45.0	22276	23889	93.2
WestCoast	241	106	3	132	44.6	20917	22833	91.6
Footscray	241	107	1	133	44.6	22759	24434	93.1
Carlton	241	105	3	133	44.2	22779	24088	94.6
PortAdelaide	165	66	2	97	40.6	14863	16985	87.5
Sydney	241	92	6	143	39.4	21401	23449	91.3
Adelaide	230	86	1	143	37.6	20199	22962	88
Melbourne	241	86	1	154	35.9	21232	24386	87.1
Brisbane	241	78	4	159	33.2	21231	25093	84.6
Richmond	241	76	2	163	32.0	20727	25673	80.7
Fremantle	187	49	0	138	26.2	15512	19909	77.9
Fitzroy	76	14	0	62	18.4	6148	9070	67.8
GoldCoast	11	2	0	9	18.2	739	1369	54

Table 5: Performance of AFL clubs away from home in the home and away rounds 1990-2011

These ladders show the large home ground advantage that exists in AFL football. Most clubs have won more matches and scored more points than their opponents at home. Only Fitzroy, Gold Coast and Richmond have won less than half their matches and points at home, while Melbourne is borderline in both. On the other hand, only Geelong and North Melbourne have won more than half their matches away, and only Geelong and Collingwood have won more points than their opponents away. The top six clubs on the away ladder are all Victorian; this is likely due to the fact that playing away within your home state is easier than playing away interstate. On the other hand three interstate sides rank in the top five in percentage of home games won, presumably for the same reason.

Because good teams win at both home and away, and poor teams lose both home and away, to properly measure home advantage one must look at the difference between home and away performance. Table 6 shows the difference in both win and point percentage between home and away, ordered in decreasing magnitude of win percentage difference. It clearly shows the effect of distance on home advantage. Six interstate sides and the only regional Victorian side are all in the top half of the table. The bottom half of the table is made up of clubs which currently share the two Melbourne grounds MCG and Docklands. Essendon appears to be an exception.

Team	Win% diff	Pt % diff	Win% Rank	Pt% Rank
Adelaide	28.5	35.8	1	1
Fremantle	28.3	26.2	2	4
WestCoast	25.3	34.1	3	2
Brisbane	24.9	27.6	4	3
PortAdelaide	23.0	25.4	5	5
Essendon	20.7	14.0	6	10
Geelong	19.9	20.5	7	6
Sydney	16.0	16.7	9	8
Richmond	16.0	12.8	8	12
Footscray	14.5	15.9	10	9
Melbourne	14.5	10.3	11	15
St.Kilda	13.9	18.7	12	7
Fitzroy	13.2	7.5	13	17
Collingwood	12.4	12.9	14	11
Hawthorn	12.2	11.9	15	13
Carlton	9.1	10.3	16	14
NthMelbourne	8.5	8.1	17	16
GoldCoast	-9.1	4.6	18	18

Table 6. Difference in home and away performance in AFL matches

Clarke and Norman (1995) make the point that any team with a real home advantage, which wins more at home than away, automatically gives their opponents the appearance of a spurious home advantage (because their opponents win less away and more at home against the team with the real home advantage). If this is taken into account a lot of the home advantage in the lower half of the table may be spurious. On the other hand, as noted above, if matches on shared grounds or other anomalies were removed from the data the home effect would presumably increase.

4 COMPUTER RATINGS.

Clarke (1988, 1993) discusses the Swinburne computer prediction program that has been published in various media outlets almost continuously since 1981. Clarke (1992) and Clarke and Clarke (1996) have shown this to be at least as accurate as the expert tipsters. For example in 2011 the computer topped the table of 14 experts and personalities in the www.footytips.com.au web site, eight ahead of the second placegetter. It finished in the top 0.3% of the 375,734 people who entered their weekly tips through this site. Only one of the 24 and 30 expert and celebrity tipsters in the Age and the Herald Sun managed to pick more winners than the computer. Since the margin predictions are essentially based on the difference in the numerical ratings the computer gives to each team playing on the given ground, the computer's prediction success suggests the ratings are reasonably valid. We use these rating as the measure of a team's current playing ability. Since the algorithm is based on an exponential smoothing system, the ratings essentially provide a weighted average of about the last eight to ten weeks form allowing for strength of opposition and including ground effect.

The algorithm used in the current study to rate teams differed slightly from that used for real time prediction. Firstly, all teams were given an initial rating of 100 in 1979 with no home advantage on any ground. Secondly there was no squashing of the ratings between seasons. Finally no practice or pre season matches were used. The algorithm was run on the complete data set from 1979 to 2011 and the actual rating of each team (including any ground effect) after each match recorded. This gave a period of 10 years for the team ratings and ground effects to build up before the start of the AFL in 1990. The 7986 ratings generated for the AFL period had a mean of 104 with standard deviation 20. This shows the first named side enjoyed an average ground advantage of four points.

In spite of the above changes, the algorithm's success rate using the pre match ratings was little worse than the original, giving some credence to the validity of the generated ratings. As these ratings are less subject to the random variations and luck of the bounce that affects wins and premiership tables, we use them in the remainder of this paper to measure the current performance level of teams.

Team	Games	Mean Rating	Mean rank	Max Rating	year,round	Max rank	Min Rating	year, round	Min rank
Essendon	512	110.5	3	165.9	2001,1	1	64.7	2006,14	7
Collingwood	510	110.1	4	165.0	2011,20	2	72.1	2000,18	3
Brisbane	507	102.2	12	163.9	2002,3	3	50.0	1994,2	16
Geelong	520	118.7	1	163.3	2008,22	4	78.9	1999,14	2
Adelaide	486	107.6	5	158.1	2006,16	5	68.6	2011,24	5
WestCoast	522	110.7	2	156.8	1991,17	6	62.2	2001,17	8
Hawthorn	504	106.9	7	153.7	1990,1	7	68.7	2006,11	4
Carlton	505	102.8	11	149.6	2000,18	8	52.0	2003,22	13
St.Kilda	507	104.5	9	148.6	2009,17	9	57.1	2001,18	10
PortAdelaide	347	103.3	10	146.3	2004,23	10	50.8	2011,21	15
NthMelbourne	509	106.9	6	136.0	1996,26	11	79.7	1990,1	1
Sydney	507	101.5	13	135.4	2008,10	12	35.0	1993,8	17
Fremantle	380	94.6	15	135.0	2006,22	13	56.8	2001,16	11
Footscray	506	105.0	8	134.6	2010,18	14	65.5	2004,19	6
Melbourne	502	97.1	14	130.9	1994,5	15	50.9	2009,3	14
Richmond	488	90.0	16	122.3	1996,21	16	56.7	2010,7	12
Fitzroy	152	73.9	17	108.6	1994,2	17	32.6	1996,6	18
GoldCoast	22	72.6	18	91.5	2011,2	18	59.6	2011,22	9

Table 7. Best and worst computer rated teams for each club.

Table 7 gives the average ratings for each club, and the teams which achieved the best and worst ratings for the club. The table is ordered on the best rating of each club, although other criteria could be used. The highest rating was achieved by the Essendon team after round 1 of 2001. This of course followed their 2000 season where they won the premiership. The Collingwood side of 2011 sits second, just before their pre final collapse. The lowest rated team is the Fitzroy team of 1996, just before they were omitted from the AFL. With the lowest average rating, the Gold Coast have struggled in their first year, with their worst rating coming at the end of the year. But compared with past teams, their worst sits half way up the table.

Paradoxically most teams achieve their absolute highest rating just before a loss, and their lowest rating just before a win. This occurs since the rating of good teams generally goes down after a loss, and that of bad teams up after a win. While higher ratings mean a team is more likely to win, it is interesting to note just how many teams fail to turn their good form into a flag. In the above table, the teams of Collingwood in 2011, Geelong in 2008 West Coast in 1991 and St Kilda in 2009 all achieved their clubs highest rating late in a year in which they won the minor premiership, but faltered at the crucial time.

5 ELITE LEVEL

How often do teams achieve a high rating? Only 15% of weekly ratings were more than 24 (one standard deviation over the mean), so we use this as a measure of excellence. Table 8 shows the number of times each club exceeded a rating of 124 along with their longest run over this figure. The top teams in this and the premiership table are the same, though clearly some clubs have been more efficient than others. St Kilda seems the unluckiest, with 95 weeks at the elite level failing to produce one premiership. On the other hand, North Melbourne have gained two premierships from just 47 weeks at the top level.

Finally we look at the length of times the elite teams remained at that level. Table 9 gives the longest sequence of continuous matches that a team remained at the elite level.

With a not out run of 114 weeks, Geelong has spent a continuous run longer than the next two teams (Essendon and Collingwood) combined. Hawthorn figure in the table three times, despite the fact their greatest period of success was before the AFL was formed. Their run of 11 at the start of the AFL was in fact the end of a run of 92 successive weeks at the elite level which began in round 19 of 1986. During the period 1982-86 they also had runs of 16, 23, 16, and 12 weeks at this level – clearly a dominant team over a long period. It is also interesting to note the strength of the competition at the end of 2011, with Hawthorn joining Geelong and Collingwood in a long run at the elite level.

Team	Number of rounds played	Number of rounds above 124	Longest run over 124
Geelong	520	203	114
WestCoast	522	145	24
Brisbane	507	113	17
Essendon	512	110	61
St.Kilda	507	95	12
PortAdelaide	347	65	9
Carlton	505	90	20
Hawthorn	504	87	18
Collingwood	510	86	47
Adelaide	486	72	15
Sydney	507	52	3
NorthMelbourne	509	47	7
Footscray	506	39	9
Melbourne	502	14	3
Fremantle	380	5	2
GoldCoast	22	0	0
Fitzroy	152	0	0
Richmond	488	0	0

Table 8 showing number of times each team had a rating over 124, and the longest run over 124.

Team	Start of run (Year, round)	Number of rounds	Included Premierships
Geelong	2007,12	114*	3
Essendon	1999,19	61	1
Collingwood	2010,5	47*	1
Geelong	1992,2	26	0
WestCoast	1991,4	24	0
Carlton	1995,14	20	1
Carlton	2000,11	20	0
Hawthorn	1991,23	18	1
Brisbane	2002,19	17	1
Hawthorn	2011,11	16*	0
Adelaide	2006,5	15	0
Essendon	1990,11	15	0
Brisbane	2001,21	14	1
Collingwood	1990,21	12	1
St.Kilda	2009,3	12	0
Hawthorn	1990,1	11*	0

* not out. Note Geelong, Collingwood and Hawthorn run still open at end of 2011. Hawthorn entered AFL with rating higher than 124.

Table 9 AFL teams that had 10 weeks or more in succession with a rating over 124.

6. CONCLUSIONS

There are many ways of ranking team performance, and most fans would argue on the merits of the well performed teams of the past. While West Coast tops the table based on premierships and finals appearances, here we have used the ratings of a successful computer prediction program to obtain more robust objective ratings. While the highest ever rating was achieved by the Essendon team on round 1 of 2001, on many measures Geelong has been the most successful team of the AFL. With the equal most premierships, the second highest number of finals appearances, the greatest percentage of wins both home and away, the highest mean rating, the highest number of rounds at an elite level and the longest continuous run at an elite level, they have an enviable record. Of course supporters of other clubs will provide counter arguments. Arguably Geelong has some way to go to match some of the dominant teams of the VFL era. Hopefully we have given some extra data for fans to consider when discussing the relevant merits of teams from different eras.

Acknowledgements

While making every effort to produce an unbiased report, we wish to acknowledge that both authors are Geelong supporters.

References

Clarke, S. R. (1988). Tinhead the tipster. *OR Insight 1(1)*: 18-20.

Clarke, S. R. (1992). Computer and human tipping of AFL football - a comparison of 1991 results. In *Mathematics and Computers in Sport*, N. de Mestre, (Ed.), Bond University: Gold Coast, Qld. pp. 81-93.

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

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

Clarke, S. R. and Clarke, R. C. (2006). A Comparison of computer and human predictions of AFL. In *Proceedings of the Eighth Australasian conference on Mathematics and Computers in Sport*, J. Hammond and N de Mestre, (Eds.), MathSport (ANZIAM): Coolongatta. pp 141-149.

Clarke, S. R. and Norman, J. M. (1995). Home ground advantage of individual clubs in English soccer. *The Statistician 44(4)*: 509-521.

Ryall, R. and Bedford, A. (2010). An optimized ratings-based model for forecasting Australian Rules football. *International Journal of Forecasting 26(3)*, 511-517.

Ryall, R. and Bedford, A. (2011). The intra-match home advantage in Australian Rules football. *Journal of Quantitative Analysis in Sports 7(2)* Article 3.

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

EVALUATING AUSTRALIAN FOOTBALL LEAGUE PLAYER CONTRIBUTIONS USING INTERACTIVE NETWORK SIMULATION

Jonathan Sargent [a] and Anthony Bedford [b]

[a,b] *School of Mathematical & Geospatial Sciences*
RMIT University

Abstract
This paper focuses on the contribution of Australian Football League (AFL) players to their team's on-field network by simulating player interactions within a chosen team list and estimating the net effect on final score margin. A Visual Basic computer program was written, firstly, to isolate the effective interactions between players from a particular team in all 2011 season matches and, secondly, to generate a symmetric interaction matrix for each match. Negative binomial distributions were fitted to each player pairing in the Geelong Football Club for the 2011 season, enabling an interactive match simulation model given the 22 chosen players. Dynamic player ratings were calculated from the simulated network using eigenvector centrality, a method that recognises and rewards interactions with more prominent players in the team network. The centrality ratings were recorded after every network simulation and then applied in final score margin predictions so that each player's match contribution—and, hence, an optimal team—could be estimated. The paper ultimately demonstrates that the presence of highly rated players, such as Geelong's Jimmy Bartel, provides the most utility within a simulated team network. It is anticipated that these findings will facilitate optimal AFL team selection and player substitutions, which are key areas of interest to coaches. Network simulations are also attractive for use within betting markets, specifically to provide information on the likelihood of a chosen AFL team list "covering the line".

Keywords: Interaction Matrix, Negative Binomial Distribution, Eigenvector Centrality, player ratings

1. INTRODUCTION

Australian Rules football, or AFL, is an invasion game played between two teams, each with 18 on-field players (and four reserves); a regular season consists of 18 teams each playing 22 matches. The dynamics of the game are similar to world football (association football or soccer), except that AFL players are permitted to use their hands to punch (handball) the ball to the advantage of a team member. The ultimate objective is to score a goal—worth six points—by kicking a ball through two upright posts at either end of the ground. Like other invasion games, scoring is the result of a series of critical events, or performance indicators, executed between the individuals involved in the contest (Nevill et al, 2002). These events are mostly discrete in nature, whether they are the number of kicks by player i or the number of times player j marks (catches) a kick from player i.

In modern sports media, player performance indicators are intensively collected and published online across an ever-increasing number of sports, both prior to and during a match. It is common for player i's indicators from a match to be weighted and linearly combined, resulting in a numerical performance appraisal, or player rating. This methodology has become a standard for many fantasy sporting leagues—that is, to calculate players' post-match ratings then proportionally adjust their (fantasy) market value according to their rating fluctuations, as determined by a moving average from past matches. A criticism of this methodology is that it is too player-centric, ignoring an important underlying concept that a team is supposed to be more than the sum of the individual players (Gould and Gatrell, 1979/80). Duch et al (2010) argue that the real measure of player performance is "hidden" in the team plays, and not derived from strictly individual events associated with player i. Moreover, in their research on football-passing patterns from EURO 2004, Lee et al (2005) measured passing between players at a *group* level rather than at an *individual* level, demonstrating how a player's passing patterns determined his location in the team network.

Discussions about network analysis commonly refer to the use of *relational data* or the interactions that relate one *agent* (player) to another and, so, preclude the properties of the individual agents themselves (Scott, 2000). The objective of this research was to move beyond such individual performance exploits, towards a measurement of each player's contribution to a dynamic system of team play. This was conceived through the identification of *link plays* within AFL matches (Sargent and Bedford, 2011), or sequences of play involving two or more players from team a where the ball's movement effectively increased scoring likelihood. Links were produced from data representing every "interaction" between the players; most games exceeded 2,500 cases. The interaction between pairs of players from team a within each link made it possible to generate an interaction matrix with which to observe player *relations*, or the number of times the ball passes from player i to player j on team a (Gould and Gatrell, 1979/80).

Symmetric interaction matrices were generated for each match played by the Geelong Football Club in 2011 and negative binomial distributions (*nbd*) fitted to each player pair in the matrix so that their interaction frequency could be simulated. Pollard et al (1977) concluded that the *nbd* is a closer fit to events resulting from groups of players, rather than from individual performances; for example, an improved fit is observed from batting partnerships in cricket, rather than from individual batsman scores. Reep and Benjamin (1968) successfully modelled effective passes in world football with *nbd*, while Pollard (1973) demonstrated how the number of touchdowns scored by a team in an American Football match closely followed the *nbd*. The *nbd* was considered to be a suitable fit to the AFL interaction data, able to simulate higher order interactions between pairs of superior players and lower order interactions between less prominent players.

After each match simulation, a rating for each player in the network was calculated using *eigenvector centrality*, a measure of the importance of a particular node (player) in a network (team)—that is, by determining the extent to which player i interacted with other central players. Centrality is a core concept in network analysis and has been applied in countless environments to determine patterns of flow, for example, infections, forwarded emails or money flowing through markets. Borgatti (2005) provides excellent definitions and applications of centrality in its various forms. The appeal of eigenvector centrality is its ability to measure the long-term influence of a node on the rest of the network, not just its immediate effect on adjacent nodes, as in *degree* centrality (Borgatti, 2005). Furthermore, a team strength index was calculated after each simulation from player centrality mean and variance, which was predictive of the team's final score margin. Through multiple iterations of the line-up and Jimmy Bartel's resulting net simulated effect on margin, the paper ultimately evaluates his contribution to a selected side.

2. METHODS

i. Player interaction

Interaction frequency between any pair of players, $[i, j]$ from team a in a match is represented by the discrete random variable, r_{ij}. Three forms of interaction were recognised from our link play data:

i) Primary interaction: efficient ball movement achieved through {Kick$_i$; Mark$_j$}, {Handball$_i$; Handball Received$_j$} or {Hit Out$_i$; Hit Out Received$_j$}[1];

ii) Secondary interaction: less efficient ball movement, namely player j gathering the ball off the ground ("Ball Get") due to an inaccurate player i event; team a retains possession of the ball;

iii) Negative interaction: inefficient ball movement where player i relinquishes possession of the ball to player k from team b ("Turnover").

The interaction methodology is similar to "r-pass movement" in world football as defined by Reep and Benjamin (1968), but is enriched by recognising the combinations of players involved in the movement.

Given the directional nature of the data within the link plays, the initial interaction matrices were asymmetric, where each A_{ij} was the frequency of player i "sending" the ball to and being "received" by player j (see points i) and ii) above). This research, however, required an *undirected* network—that is, any and all relations between players regardless of the directional flow (Scott, 2000). A *directed* network would be preferred if we were interested in a player's send/receive ratio. For example, because he is mostly attempting to score, a forward would receive the ball from teammates more than he would send the ball. The undirected

[1] A "Hit Out" is similar to a jump start in basketball, except the competing players must "tap" the ball down to the advantage of a team member.

network required each matrix to be symmetrised, using $r_{ij} = r_{ji} = A_{ij} + A_{ji}$, $(i, j = 1,\ldots,22)$. Frequency distributions could then be calculated for each $[i, j]$ in each of Geelong's 25 matches (22 regular season games and three finals matches). Geelong fielded 34 players throughout the season, so a total of (34 x (34-1))/2 = 561 distributions were computed. In this calculation, the subtraction of 1 removed player i's interaction with himself, and the divisor of 2 halved the distributions to be calculated because $r_{ij} = r_{ji}$. Figure 1 displays the observed interaction, $f(r)$, between Geelong's Jimmy Bartel and Andrew Mackie for all 2011 season matches. This player pair was more likely to interact between one and six times in a match than not at all. The maximum number of interactions measured in the season between any pairing from the team was eight.

Figure 1. Frequency distribution of [Bartel, Mackie] interactions

ii. Interaction simulation

If the average rate of discrete events that occur between two players within an AFL match remained constant over its course, the events could be described with a Poisson distribution. However, interaction rates between any $[i, j]$ are stochastic, depending on factors such as the position of the two players, their skill levels and the defensive quality of the opposition. For this reason, the negative binomial distribution (*nbd*) was deemed more appropriate than Poisson. Although the performances of individual players do not give close fits to the *nbd*, the fit improves as more players become involved (Pollard et al, 1977). From the negative binomial distribution, the probability of r interactions for each $[i, j]$ is:

$$P(r) = \binom{k+r-1}{k-1} p^k q^r \text{, } r = 0,\ldots,8 \qquad (1)$$

where $k > 0$, $0 < p < 1$ and $q = 1 - p$.
The parameters, k (the threshold number of successes) and p (the probability of a success) were estimated so as to minimize the Pearson's chi-squared statistic, χ^2 for each $[i, j]$, by using the observed (O) and expected (E) probabilities derived from Equation (1), or:

$$\min \chi^2 = \sum_{r=0}^{8} \frac{(O-E)^2}{E} \qquad (2)$$

$$\text{s.t. } 0 < p < 1 \text{ and } k > 0.$$

where r is the number of failures (interactions).
Fitting *nbd* to various sports, Pollard et al (1977) estimated k and p by a method of moments, so:

$$k = m^2/(s^2 - m),\ p = m/s^2 \qquad (3)$$

where m is the sample mean and s^2 is the sample variance. We concluded that Equation (2) was a more adequate fit to the interaction data, providing lower χ^2 values for the majority of Geelong's $[i, j]$. The [Bartel, Mackie] example is displayed in Table 1 where k and p in each $P(r)_1$ were estimated using Equation (2) and in each $P(r)_2$ using Equations (3).

r	$f(r)$	$P(r)_1$	$P(r)_2$
0	0.3333	0.3333	0.2897
1	0.1905	0.2222	0.2675
2	0.1429	0.1481	0.1853
3	0.1429	0.0988	0.1141
4	0.0952	0.0658	0.0659
5	0.0476	0.0439	0.0365
6	0.0476	0.0293	0.0197
7	0.0000	0.0195	0.0104
8	0.0000	0.0130	0.0054
χ^2		0.0819	0.1178

Table 1. Probabilities and χ^2 values for [Bartel, Mackie]

A Visual Basic module was written to fit the optimized *nbd* to all combinations of players in the Geelong club and to simulate the players' interactions for any chosen team list in the 22 x 22 team matrix. The initial routine produced a random probability, $u \sim U(0,1)$ for each $[i, j]$ in the match, with r_{ij} determined by the cumulative distribution function:

$$F(r) = P(R \le r) \qquad (4)$$

where R represents the cumulative probability. For example, a randomly generated probability of $u = 0.3$ would produce $r_{[\text{Bartel, Mackie}]} = 0$ as $u < P(R \le 1) = 0.0000 + 0.3333$ (see Table 1). For each simulation, all (22 x (22 - 1))/2 = 231 elements of the interaction matrix assumed a value for r as determined by u and Equation (4), enabling calculation of player ratings from the simulated matrix.

iii. Player ratings

Measuring a player's net contribution to a match in any team sport is an ambiguous task, in particular for the AFL, because 36 players compete on the field at any single moment. The different positional duties performed by each player add to the complexity: defenders prevent goals; forwards kick/create goals; and midfielders obtain and retain possession of the ball to increase the chance of their team scoring. A network algorithm was introduced for rating purposes to better understand the causality of player i's performance with respect to that of his teammates.

Centrality is one of the most widely studied concepts in network analysis and allows implicit assumptions about the prominence of an individual in a network (Lusher et al, 2010). A specific type, *eigenvector* centrality, was trialled as a valid player-rating model, under the assumption that the higher a player's centrality in the Geelong network, the greater his interaction with other players. The eigenvector centrality rating, e, for player i, was measured using:

$$e_i = \frac{1}{\lambda}\sum_j r_{ij}x_j \qquad (5)$$

expressed in matrix form as: $Ax = \lambda x$, where x is the corresponding eigenvector from our interaction matrix, A, and the eigenvalue, λ, was solved using an automated power method. Following n multiplications of A and x, the point at which λ_{n-1} and λ_n converged prompted calculation (Equation (5)) of the ratings for all players within the actual or simulated interaction matrix.

The simulated network and corresponding ratings detailed in this paper provided a pragmatic framework for estimating player i's utility within a selected side. An important step in this procedure was calculating team a's network "strength", π, after each match, by:

$$\pi_a = \frac{1}{n}\sum_{i=1}^{n} e_i \, , \; i = 1,\ldots,22 \qquad (6)$$

where each e is derived from Equation (5). We compared Geelong's 25 network indices from Equation (6) with each match's final score margin and discovered a linear regression line effectively approximates the margin (R^2 = 0.5302) (see Figure 2i)). In practical terms, a team increases its likelihood of winning if more players force themselves to be central in the match network. This is analogous to the finding that soccer teams, skilful enough to retain possession for longer periods than their opposition, have a greater chance of scoring (Hughes and Franks, 2005).

To validate the centrality ratings, an "individual" rating equation, Y_i, was developed, ignoring network methodology and focussing solely on player i's post-match performance indicator totals—the same four indicators (m) as in the primary interaction data (kick; mark; handball; handball received). The equation was of the form:

$$Y_i = b_o + \sum_{m=1}^{4} b_m X_m \qquad (7)$$

where X_m is the frequency of performance indicator m for player i, b_m are weights and b_o is the intercept. The weights were optimized to maximize the linear relationship between the mean ratings and final score margin in each Geelong match. Substituting Y for e in Equation (6) produced a comparable measure of team strength for the individual ratings. Figure 2ii) confirms team strength was not as accomplished at predicting score margin when each player was assessed individually (R^2=0.2837), rather than as an agent within a team's network.

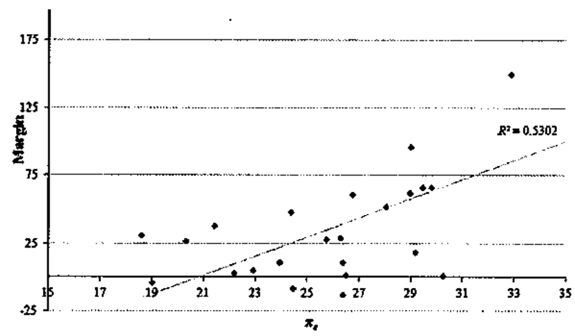


Figure 2i). Relationship between Geelong's mean network rating and final score margin

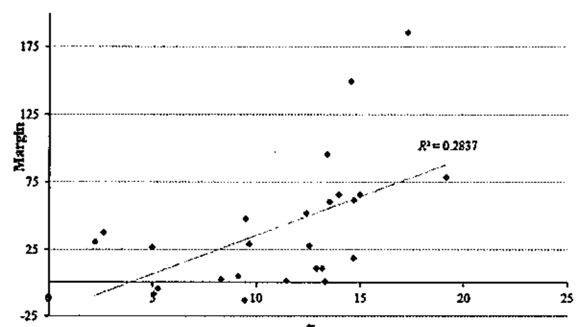


Figure 2ii). Relationship between Geelong's mean individual rating and final score margin

3. RESULTS

Before investigating player effects within the network, we performed a preliminary examination on our simulator, testing the hypothesis of similar means between the observed and simulated total interactions from Geelong's 22 regular season matches. Opposition effect was ignored for this stage of the research. One hundred interaction simulations were run on each round's interaction totals, Σr_i, (i = 1,..,22), and the mean and standard deviation of each distribution compared with the total observed interactions in each match. Figure 3 reveals a satisfactory fit for the model, with no significant difference between the simulated and observed series means (p = 0.764, α = 0.05). Moreover, the majority of observed totals fell within 95% confidence intervals associated with each simulated match mean. Match 13 was considered an anomaly in the series—Geelong fielded their weakest side for the season, as acknowledged by the simulator, but managed to achieve almost 600 interactions and to win by 52 points, most likely due to their home-ground dominance. The outlier at Match 18 was Geelong winning by 186 points—the second-highest margin in AFL history—yet the simulator acknowledged the strength of this side, offering the largest simulated interaction mean of all matches (Σr_i = 654). The overall fit gave us confidence to proceed to analysis of individual player effects.

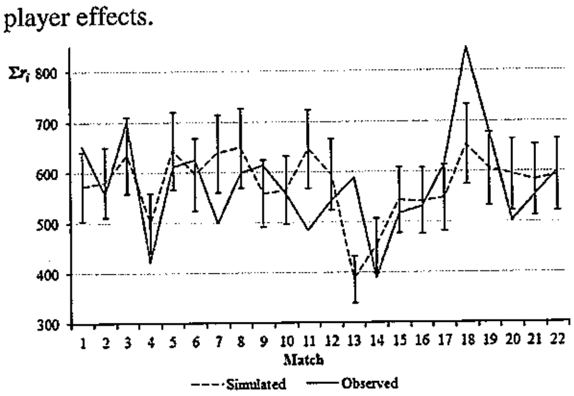


Figure 3. Simulated and observed interaction totals for Geelong matches

A case study was undertaken on Geelong's 2011 grand final team list, beginning with one thousand network simulations. Using the regression line in Figure 2i), final score margins were predicted and logged after each simulation. The black curve in Figure 4 represents the normal distribution ($\bar{X}_1$ = 47.03, σ_1 = 14.31) of predicted margins given Geelong's actual grand final network. Geelong won the game by 38 points, which reflects the model's predictive properties. Another one thousand simulations were run on the same side, but we replaced Bartel with a player of lesser skill, Shannon Byrnes. The light grey curve in Figure 4 represents the normal distribution ($\bar{X}_2$= 31.96, σ_2 = 13.15) of margins after Byrnes replaced Bartel in the side.

Interpretation of this result is important; we concluded that, given his replacement (Byrnes), Bartel's estimated net contribution to the <u>selected team</u> was $\bar{X}_1$ - $\bar{X}_2$ = 15.07 points. Stressing the selected side was necessary as it could be hypothesised that Byrnes replacing Bartel in a stronger side may have less impact on margin due to the contribution of the other high-calibre players. To conceptualise the importance of selecting the best replacement player, we ran a third iteration in which we replaced Bartel with Darren Milburn—a highly regarded player, but not as skilful as Bartel—and again ran one thousand simulations. The normal distribution ($\bar{X}_3$= 42.85, σ_3 = 14.87) is represented by the dark grey curve in Figure 4, from which we concluded that, given his replacement (Milburn), Bartel's estimated net contribution to the selected team was $\bar{X}_1$ - $\bar{X}_3$ = 4.18 points. The difference between the mean of the Byrnes and Milburn distributions ($\bar{X}_2$ - $\bar{X}_3$ = -10.89) implied a coach would be more inclined to replace Bartel with Milburn in that side because the negative effect on margin is reduced. It is logical that a player may be selected on grounds other than his net effect on margin; for example, Byrnes's style of play may be more suited than Milburn's to the game-day conditions, but this is outside the concerns of this paper.

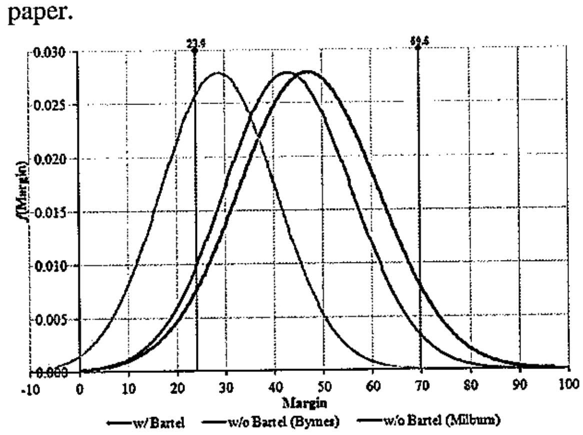


Figure 4. Margin distributions with and without Bartel

4. DISCUSSION

If a prominent player is removed from the network, remaining r_{ij} distributions are not recalculated—that is, we assume teammates do not improve their performance to cover the absence of the excluded player. This phenomenon of players exceeding expectation will be explored further in ongoing research. Furthermore, this paper has not considered the presence of covariance between any r_{ij}. The initial stages of this research governed that each r_{ij} is independent, even though degrees of interaction covariance between sets of $[i, j]$ are almost certain. The thousands of $[i, j]$ permutations and covariance between each would command a separate research paper.

Ongoing research will also focus on improving the predictive power of the networks by weighting the three forms of player interactions in Section 2i with respect to the levels of efficiency, scoring capacity and ground and opponent effects.

5. CONCLUSION

Player-based statistical analysis is as important in today's sporting environments as ever before, with coaches continuously searching for the right mix of players to include in a team. In the AFL, the decision to include in a team one player over another can have serious repercussions on the outcome of the game. We developed a model to assist in such selection decisions by simulating different players' interactions with one another and by measuring the effect of such networks on final score margin. Negative binomial distributions were fitted to all pairs of players within a side so that interactions between players could be simulated prior to a match. It was discovered that the strength of the Geelong team's networks was predictive of its final score margin; therefore, it was possible to measure the contribution any player could make to the final margin. Hence, when a team's line-up is revealed, so too is the likelihood of the team winning. From a pre-match betting perspective, it is possible to calculate the odds of the selected team "covering the line". It is anticipated that an in-play model will add further value because coaches and punters can make informed decisions with knowledge of live match scenarios.

References

Borgatti, S.P. (2005) Centrality and network flow. Social Networks, 27, 55-71.

Duch, J., Waitzman, J. S. and Amaral, L. A. N. (2010). Quantifying the performance of individual players in a team activity. PLoS One, 5, 1-7.

Gould, P. & Gatrell, A. (1979/80). A structural analysis of a game: the Liverpool v Manchester United Cup final. Social Networks, 2, 253-273.

Hughes, M. & Franks, I. (2005). Analysis of passing sequences, shots and goals in soccer. Journal of Sports Sciences, 23(5), 509-514.

Lee, J., Borgatti, S. P., Molina, J. L. and Guervos, J. J. M. (2005). Who passes to whom: analysis of optimal network structure in soccer matches. Proceedings of the Twenty-fifth International Sunbelt Social Network Conference, Redondo Beach, CA.

Lusher, D., Robins, G. and Kremer, P. (2010). The application of social network analysis to team sports. Measurement in Physical Education and Exercise Science, 14, 211-224.

Nevill, A.M., Atkinson, G., Hughes, M. D. and Cooper, S-M. (2002) Statistical methods for analysing discrete and categorical data recorded in performance analysis. Journal of Sports Sciences, 20, 829-844.

Pollard, R. (1973). Collegiate football scores and the negative binomial distribution. Journal of the American Statistical Association, 68, 342, 351-352.

Pollard, R., Benjamin, B. & Reep, C. (1977). Sport and the negative binomial distribution. Optimal strategies in sport (S.P. Ladany & R. E. Machol eds.) pp. 188-195. Cleveland: North-Holland Publishing Company.

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

Sargent, J. & Bedford, A. (2011). Influence and prestige: rating Australian Football League player performance with network analysis. Proceedings of the 3rd IMA International Conference on Mathematics in Sport. Manchester, UK. (Compact disc).

Scott, J. (2000) Social Network Analysis: A Handbook. Sage Publications Ltd.

PREDICTIVE SUCCESS OF OFFICIAL SPORTS RATING SYSTEMS IN INTERNATIONAL COMPETITION

Raymond T. Stefani

California State University, Long Beach, USA
Corresponding author: Raystefani@aol.com

Abstract

At Mathsport 2010 in Darwin, 160 officially-recognized world sports were surveyed, including combat, independent and objects sports. The organizing federations published 100 official rating systems, falling into three categories: subjective, self-adjustive and point-accumulation. That study is extended by analysing the predictive success of five self-adjustive systems and seven point-accumulation systems applied to object sports (arguably among the hardest sports to predict the winner). For selected World Cup, Commonwealth Games and tennis Grand Slam competitions, the percent of matches was tabulated in which the higher ranked competitor won. Ties were considered half correct where ties are infrequent in the group phase and fully-incorrect in football in the group phase. To compare football with the other sports, predictive success was also tabulated as percentage above random chance. The top four systems in terms of accuracy were self-adjustive (women's netball, men's rugby, women's football, and men's cricket), scoring 40-29% above random chance (90-79% without ties). The IFNA netball system had 90% overall success in two competitions while the IRB rugby system had 90% success in the last rugby WC, surprisingly high values. Surprisingly also, in the knockout phases of the last FIFA football WCs, the higher ranked women's teams won 81% of the matches while men won 88%, in a low scoring sport. The fifth self-adjustive system for men's football was near the bottom at 20% above random chance but that system resembles a point-accumulation system more than a self-adjustive system. The 7 point-accumulation systems scored from 29-19% above random chance (79-69% without ties): men's and women's tennis, men's and women's volleyball, men's curling, men's ice hockey and men's basketball. Corrected for competitive balance, the top four self-adjustve systems at 33-27% above random chance (83%-77% without ties) and the 7 point-accumulation systems at 23-19% above random chance (73%-69% without ties) separated more obviously.

Keywords: predictions, rating systems, competitive balance, world cup, Elo, probit

1. INTRODUCTION

In Stefani (2010, 2011), a sport is defined as a competition using established rules for determining the winner. Sports so defined fall into two classes, mind sports and physical sports. In a mind sport, a surrogate (human, mechanical or computer) can make a play for the competitor as in chess and bridge. No physical action is need. In a physical sport, the competitor must make each play, requiring physical prowess as in running and swimming. What I call a physical sport follows the more widely used dictionary definition of a sport; however I follow the lead of the IOC which lists chess and bridge among "recognized sports".

In turn, these sports fall into three categories. First, in a combat sport, competitors are in direct contact: the goal is to control the opponent as in wrestling and boxing. Second, in an independent sport no significant contact is allowed as in running and swimming: the goal is for the competitor to control his or her own self. Third, in an object sport, the competitors interact indirectly as in soccer, chess and rugby: the goal is to control an object. The sports and recognizing agencies are tabulated in Stefani (2010, 2011).

The rating systems also fall into three categories, called subjective, accumulative and adjustive herein. First, a subjective poll of authorities is used in a few combat sports. Seconds, points are tabulated over some window by point accumulation systems, simply termed accumulative. Such systems may be described as follows where r represents an accumulative rating.

$$r = \Sigma f_i\,(results,\ importance,\ ageing) \quad (1)$$

Summation is over all of the competitors past performances in the window indexed by i, which may cover one or more years. Each f_i is non-negative so the sequence of running sums is non decreasing, hence the term "accumulative". The function f_i converts each result to points which are modified by the importance of the event and then aged depending on when the event happened. For a four-year window fro example, each year's results could be weighted 100%, 75%, 50% and 25% in order from the most recent year.

The third type of rating system self adjusts each previous or old value of r creating a new value after each new result, hence the term "adjustive".

$$r(new) = r(old) + K\,[\,result - prediction\,] \quad (2)$$

The prediction is usually based the difference of the old ratings of the competitors. An adjustive rating system is inherently more predictive than an accumulative system since the ratings self adjust to to correct for any past predictive errors.

Table 1 summarizes the types of sports versus the types of rating system, with a few updates from the tables in Stefani (2010).

Table 1 Types of Sports Rating Systems

		Type of Sports Rating System			
Sport	Number	None	Subj.	Accum.	Adj.
Combat	18	12	2	3	1
Indep.	74	18	0	53	3
Object	67	30	0	28	9
Total	159	60	2	84	13

The sports in Table 1 are those organized by federations recognized by the IOC, by Sport Accord, an international agency, and by a few additional federations from Wikipedia under "List of international sport federations". See Stefani (2010, 2011) for tables and specifics. Of the 159 sports, 18 are combat, 74 are independent and 67 are object. Of the 99 rating systems, 2 are subjective, 84 are accumulative and 13 are adjustive. Subjective systems are preferred by combat sports, accumulative systems are most preferred by independent sports and adjustive systems are most preferred by object sports. The bulk of the ratings are accumulative since those systems are simple and encourage competitors to enter as many competitions as possible. Accumulative systems satisfy the needs of tournament directors and are easily understood by the pubic; hence, accumulative systems are preferred by independent sports wherein results are separable for each competitor and a simple awarding of points or choosing the best previous result accurately separates competitors. The hardest sports to rate are the object sports wherein the each result of one competitor directly affects that of the other. The rigor of an adjustive system is preferred for such systems wherein ratings seek a level depending on past opponent ratings and results.

The remainder of the paper is organized as follows. In Section 2, some sports and rating systems are chosen for a study of predictive success. Section 2 also covers the handling of ties and evaluation of competitive balance. Section 3 compares and contrasts the raw results and then the results corrected for competitive balance. Section 4 contains conclusions.

2. METHODS

Sports and Rating Systems Chosen

An objective study of the predictive success of ratings applied to combat sports would be confounded by the fact that winners are generally determined subjectively. Combat sports were not studied. Independent sports such as swimming and running create a rank order of finish. Predictive success would be measured by positional differences between per-race rank order and competition rank order, which would be hard to compare to head-to-head contests. Further, such differences might be due more to inconsistent performance of the competitors than due to the rating system. Independent sports were not studied. Object sports were chosen since sports like basketball, soccer and rugby provide a challenge to rating and predicting.

Table 2 shows the 12 competitions in the nine sports used in this study. Five have adjustive rating systems while the remaining seven have accumulative systems. Three sports involve competition for both men and women.

Table 2 Competitions and Rating Systems Chosen for this Study

Self Adjustive (5)		Point Accumulation (7)	
Men	Women	Men	Women
Cricket		Basketball	
Football	Football	Curling	
Rugby	Netball	Ice Hockey	
		Tennis	Tennis
		Volleyball	Volleyball

Two factors limited the number of events available to evaluate predictive success. The first factor was the year the rating systems launched and the second factor was whether archived ratings subsequently existed prior to each event. The events and event winners are shown in Table 3. For tennis, 16 Grand Slam events were used for both men and women. In five of the other 10 competitions, two events were available and in five one event was available. Australia and New Zealand were each double winners.

Table 3 Events and Winners

Sport	*Fed.*	*Year (Winner)*
Basketball	FIBA	2010 (USA)
Curling	WCF	2011 (Can.)
Cricket	ICC	2007 (Aust) 2011 (India)
Football(W)	FIFA	2007(Germ.) 2011(Japan)
Football (M)	FIFA	2010(Spain)
Ice Hockey	IIHF	2010(Czech.) 2011 (Fin.)
Netball(W)	IFNA	2010(N. Zeal.) 2011(Aust)
Rugby	IRB	2007(S. Afr.) 2011(N. Zeal.)
Tennis (W)	WTA	16 Grand Slams, 2006-2012
Tennis (M)	AAP	16 Grand Slams, 2006-2012
Volleyball(W)	FIVB	2011 (Italy)
Volleyball(M)	FIVB	2011 (Russia)

Table 4 summarizes the properties of four of the five adjustive systems from Table 2. These four differ as to quantifying the result, predictive method and on the *K* value that implement (2). More details are in Stefani (2010,2011). The FIFA women's football (soccer) system employs the same basic Elo system as is used in chess, except that all wins are not valued 1 and all losses are not valued 0 for the result, as in chess. The margin of victory, MOV, produces a 0-1 scale. The prediction depends on the rating difference between the team with rating *r* and the opponent with rating r_O, so that
$d = r(old) - r_O(old)$. Home advantage is added. The Elo probability of *r* winning is
$1/(1+10^{-d/400})$. The adjustment *K* factor depends on importance.

Table 4 Four of the Adjustive Systems

Sport (Federat.)	Adj. Syst.	Result	Prediction	K	Other
Football (W, FIFA)	Elo Logit	0-1 based on MOV	P(d) Elo	Import.	Home Adv.
Rugby (IRB)	Probit	(L=-1, T=0, W=1)	d/10 with limiting	Import. MOV	Home Adv.
Cricket (ICC) Netball (W, IFNA) (both by David Kendix)	Probit	(L=-1, T=0, W=1)	d/50 with limiting	50/n Import.	

The systems used by rugby, cricket and netball define the result with 1 being a win, 0 being a tie and -1 being a loss. For rugby, the prediction is given by *d/10* with limiting to the range -1 to 1 and with a home advantage included. *K* depends on match importance and on MOV. David Kendix created the ICC cricket and IFNA netball systems. In a private correspondence, Kendix indicated that he first was asked by the ICC to create a system. He gave them a questionnaire which resulted in the system. The ICC did not wish to use home advantage and MOV. Later, the IFNA wanted the same system. The prediction is given by *d/50* with limiting to the range -1 to 1. *K* depends on importance and the number of games previously played. See the References section for links to Kendix' ICC and IFNA systems.
The FIFA men's football (soccer) system is adjustive in that ratings can move in any direction via an average. Any average can be posed in the

form of (2). Match points start with a scale of win-lose-tie of (3,1,0 or 2,1 for OT) times [200 – opponent rank]/100] which is in the range (.5-2) times the importance of the competition on a (1-4) scale times a continental factor on a scale of (.85-1) times 100 for a range of 0-2400. Points are summed for each year, divided by the number of games for each year and then age-weighted over four years with the current year counting the most. If teams play about the same number of matches, then division by the number of games is just a scaling factor and rank order would be the same as for an accumulative system.

The seven accumulative systems have various point scales based on position of finish in each event and the importance of the event. The number of years varies. Points are accumulated for each year and then age-weighted. The sports competitions (years) are FINA basketball (8), WCF curling (6), IIHF ice hockey (4), FIVB men's and women's volleyball (4), ATM men's tennis (1) and WTA women's tennis (1).

Ties

If a match can be tied, that presents a problem in tabulating the predictive success of the rating system in that match and then fairly comparing that system with systems in sports where there cannot be a tie. All matches in tennis are knockout matches so that there cannot be any ties. The other sports have group phases, where ties can happen in a small number of sports followed by knockout phases where ties cannot happen in any sport. A tie can happen, but is infrequent, in the group phase of rugby and cricket. Ties can happen in almost 1/3 of the group phase matches in football. Ties are so infrequent in rugby and cricket that bookmakers seldom offer odds for a tie and if one happens, half the money that would have been won is returned. Following that, ties were counted as half correct and half incorrect in those sports. In football, bookmakers do offer odds for a tie. If a win is bet but a ties results, all money is lost. Following that, ties in football were considered incorrect predictions. To compare rating systems fairly, predictions were adjusted for random chance and then ranked via predictive success above random chance. Accordingly, 33% was subtracted for the accuracy in football group phase matches since random selection of win-tie-loss would result in 33% accuracy. In all other cases, 50% was subtracted as the random chance correction.

Competitive Balance

The spread of talent in the 12 competitions was not necessarily equal. To measure competitive balance, let

$$\text{Comp. balance} = 100\,[\text{entrants/rated}] \quad (3)$$

If a small fraction, say the top 25% of rated teams makes the world cup field, then there would be reasonable balance among teams and that would challenge a rating system. If a high fraction of rated teams are present, then the skill differences among teams would be greater and early matches relatively easier to predict. In what follows, systems are ranked by success above random chance in Table 5, regardless of competitive balance. In Table 6, adjustments are made for competitive balance.

3. RESULTS

Ranked by predictive success above random chance, RC, the top four systems are adjustive in Table 5. The predictor-corrector form of (2) causes ratings to follow adjust in response to actual results over time and logically become more accurate. The bottom eight systems include the seven accumulative systems and men's football, which is effectively accumulative if the rated teams have played abut the same number of matches. It is logical that the FIFA men's football system would act similarly to the accumulative systems.

The competitive balance measure is well above the others for netball, cricket and both tennis competitions, and will be compensated for shortly. The oddly high value of 400% in tennis for (3) results in following only the 32 seeds out of the 128 entrants on each of men's and women's Grand Slam singles events, that is , 100*128/32 = 400%. The seed ranking generally follows the WTA and ATP tour ranking with some small adjustments made by tour organizers to create popular matchups. It is reasonable to assumed that the success of the top seeds mirror the success of the higher ranked players.

There are some surprises in Table 5. Except for football, accuracy in the knockout phase is about 12% less than in the group phase as the remaining competitors become more closely ranked.

Table 5 Predictive Success of the Selected Recognized International Sports Rating Systems

Sport	Federation (Rating System)	Games (WC, GS.)	Comp Bal	Higher Ranked Team Won (%)			
				Grp.	Knock out	All	All-Rand Chance
Netball(W)	IFNA (Adj)	86(2)	56%	94	81	90	40
Rugby	IRB (Adj)	96(2)	21%	85	75	83	33
Football(W)	FIFA (Adj)	64(2)	13%	67	81	70	33
Cricket	ICC (Adj)	96(2)	127%	80	69	79	29
Tennis (M)	ATP (Acc.)	1385(16)	400%		79	79	29
Tennis (W)	WTA(Acc.)	1400(16)	400%		77	77	27
Volleyb(M)	FIVB(Acc.)	66(1)	9%	71		71	21
Volleyb(W)	FIVB(Acc.)	66(1)	9%	71		71	21
Curling	WCF (Acc.)	71(1)	27%	71	67	70	20
Football(M)	FIFA(Adj.)	64(1)	15%	48	88	58	20
Ice Hockey	IIHF(Acc.)	116(2)	25%	72	56	70	20
Basketball	FIBA (Acc.)	80(1)	32%	70	65	69	19

Table 6 Predictive Success Corrected for Competitive Balance

Sport	Rating System	Higher Ranked Team Won (%) Corrected for Comp. Balance	
		All	All-Rand Chance
Netball(W)	Adjustive	83	33
Rugby	Adjustive (Uses HA and MOV)	83	33
Football(W)	Adjustive(Uses HA and MOV)	70	33
Cricket	Adjustive	77	27
Tennis (M)	Accumulative	73	23
Volleyball(M)	Accumulative	71	21
Volleyball(W)	Accumulative	71	21
Curling	Accumulative	70	20
Football(M)	Adjustive	58	20
Ice Hockey	Accumulative	70	20
Tennis (W)	Accumulative	69	19
Basketball	Accumulative	69	19

In football, knockout phase accuracy rose to 81% for women's football and to 88% for men's football. Considering that only 2.5 goals per game are scored, any bad-luck goal on a penalty kick or own goal makes up 40% of the game score. The higher ranked team did amazingly well. The seven accumulative systems and men's football scored between 69% and 79% accuracy, which is surprising good considering the simplicity of those systems.

David Kendix' IFNA netball system was correct in 90% of the matches for two events. That sort of accuracy for world class competition was not expected. When the value for the 2011 WC showed 90% accuracy, the 2010 Commonwealth Games were examined to make sure the 90% figure was not a fluke. The accuracy for the Commonwealth Games was also 90%. Another surprise was that the accuracy for the 2011 rugby WC was 90%.

Compensating for Competitive Balance

In Norton (2004), a common rating system was applied to compare the accuracy of netball predictions (80%) to Australian Rules Football predictions (70%) implying that netball has an inherent 10% added accuracy due to less competitive balance. In Norton (2004), a common system compared Australian women's WNBL basketball accuracy (76%) to men's NBL (70%) for a 6% add-on in predictive success for a woman's net sport. I have compiled nine years of comparisons looking at the higher seed in the USA national NCAA basketball tournament with the higher seeded women's teams winning 79% compared to men's teams at 73%, a 6% add-on for women. Based on those three comparisons, net sports for women are 7% more predictable using a common paradigm so that 7% was subtracted from the 90% netball accuracy creating a corrected value of 83% (33% above RC). For tennis, 32 players remain starting with round three of a Grand Slam tournament. Given that there are 32 seeds, most of the remaining players are seeded (rated) creating an approximate competitive balance of 100%, an improvement over 400%. Finally, cricket matches among non-seeded teams were not used which made the effective competitive balance figure 100%, The results are in Figure 6.

4. CONCLUSIONS

Corrected for competitive balance, the top three systems in Table 3 have identical success vs. RC, 33%, equal to 83% accuracy without ties, For top level competition, that is surprisingly good accuracy. The top three systems are the IFNA netball system, the IRB rugby system and the FIFA Elo women's football system. It should be noted that the Elo women's football system outperformed the men's system in the group phase by 19%, hence it would be a better tool for seeding WC competition. FIFA would be well advised to use the Elo system for both genders. The rugby and women's football systems use home advantage and margin of victory.

At the lower end of the scale are the seven accumulative systems and the FIFA men's football system, within four percent (19-23% above RC or 69-73% absent ties). That is remarkable accuracy for such simple systems. The cricket system lies between the two clusters of systems, being an adjustive system with 77% success (27% above RC) in a challenging sport. The top four adjustive systems (created as predictor-correctors) are, on average, 11% more successful than other eight.

5. REFERENCES

Kendix, D. (2012) IFNA rating system: http://www.netball.org/ifna.aspx?id=94.

Kendix, D. (2012a) ICC rating system: http://static.icc-cricket.com/ugc/documents/DOC_BA4FD677CDC6239 1608CA0F6662C7DEB_1338990291442_280.pps.

Norton, P. (2004). Is There a Difference in the Predictability of Men's and Women's Basketball Matches? *Proceedings of the Seventh Australasian Conference on Mathematics and Computers in Sports, Massey University, New Zealand.*

Stefani, R. (1999) A Taxonomy of Sports Rating Systems, *IEEE Trans, on Systems Man and Cybernetics, Part A*, 29, 116-120.

Stefani, R. (2010). A World of Sports and Rating Systems, *Proceedings of the Tenth Australasian Conference on Mathematics and Computers in Sports, 5-7 July, Darwin Australia, Mathsport 2010.*

Stefani, R. (2011) The Methodology of Officially Recognized Sports Rating Systems, *Journal of Quantitative Analysis in Sports*, 7(4), article 10.

AFL CLUBS AS COMMUNITIES OF PRACTICE

Kelly Foreman [a b d]

[a] *University of South Australia*
[b] *Reconfigurable Computing Lab*
[d] *Corresponding author: Kelly.Foreman@unisa.edu.au*

Abstract

A community of practice (COP) is a group "of people informally bound together by shared expertise and passion for a joint enterprise" (Wenger and Snyder 2000, p.139). Participants of the community of practice share their knowledge and experiences with the others and work together to come up with new solutions to problems and adaptable knowledge. Knowledge creation, sharing and transfer when fostered through the right channels is a source of power and competitive advantage to organisations and COP's assist in these group learning situations. All of which can be applied to a group of coaches in Australian Rules Football including activities such as problem solving, requests for information, seeking experience, reusing assets, coordination and synergy, discussion off developments and documentation of project.

This paper argues that there is a clear need for communities of practice with AFL clubs. Information as part of the community can be leveraged to provide a competitive advantage and improve coaching and individual player performance when coaching staff and players are active as members of the community of practice that sees knowledge sharing and transfer amongst the football department.

Keywords: AFL, Communities of Practice, Knowledge Sharing, Knowledge Transfer

1. INTRODUCTION

Australian Football League (AFL) clubs have a significant problem with the overwhelming amount of information they receive live during the game and post-game which is often required to be shared amongst several members of a large football department. Most clubs have between 25-30 staff in the football department alone whilst the majority of clubs employ over 100 staff in total. Clubs who own 42 GPS units (one for each player) have data live during the game coming in for analysis every few seconds, champion data statistics, video, wellness information and surveys, injury management and rehabilitation information, opposition and trends analysis and general reporting performed by the football department on its players. In the majority of clubs this information is stored in spreadsheets, flat files, email, handwritten documents and word files in various locations throughout the organisation by different staff members. An example of this can be seen within the fitness department of a football club with the following case: A players GPS numbers have been low over an unspecified period of time during games and training, the head fitness coach speaks with the player and asks if an injury has occurred or if there is another issue which is why he hasn't been performing. The player informs the fitness coach that he is not injured and there are no other issues and that he feels fine, the fitness coach just assumes that the player is not putting in the effort or his game and performances are not peaking at the right time. Rather than sharing this information with the welfare manager, line coach, senior coach, physio or doctors to consult with them to determine what further steps can be taken he keeps the information within the fitness department as they are the prime users of fitness related GPS data.

Similarly to the football example provided in the previous section the following scenario/example applies to business more specifically engineering – A group of employees within an engineering department are working on a project which involves developing a new method to reduce Australia's carbon footprint rather than collectively working together they are working individually. When a community of practice is applied to this scenario it would see the engineers share knowledge regarding their process and the techniques they have applied so far, design methodologies, successes and failures, test plans and potential ideas that need further development. The knowledge sharing and transfer that occurs as part of the community of practice due to the one common goal would see everyone benefiting from the community and the ability to work together and share knowledge and data within the project rather than trying to resolve the issue as an individual who does not have sole knowledge or expertise in the area. Although communities of practice have not yet been applied to an AFL football department there is the potential that would see it benefit clubs.

2. LITERATURE REVIEW

Introduction to Knowledge Management

Scarbrough, Swan & Preston (1999, p.669) define knowledge management as "any process or practice of creating, acquiring, capturing, sharing and knowledge, wherever it resides to enhance learning and performance in organisations". This is the definition that will be used throughout this research and the three components of knowledge management - knowledge creation, capture and acquisition and shared knowledge will be discussed.

Alfred Marshall once said "knowledge is our most important engine of production" (Sallis and Jones 2002, p.1). There are two types of knowledge tacit and explicit (Collins 2010). Magalhaes (2004, p.79) defines tacit knowledge as "personal and context-specific; it is also hard or sometimes impossible to articulate the language...it is not easy to explain what you actually do when undertaking these activities, although they can be taught or explained by means of analogy or metaphors". Tacit knowledge is knowledge gained through a personal experience whilst explicit knowledge is rational and sequential knowledge that is easy to pick up and understand with no previous experience or context. Magalhaes defines this as "knowledge which is modifiable and transmittable in formal language" (Magalhaes 2004, p.79). Knowledge creation is the process of sharing tacit knowledge as most members of an organisation, this knowledge is then converted into explicit knowledge as it has been expressed or taught through the use of metaphors or analogies (Von Krough et al. 2000).

Awad & Ghaziri (2004, 147) define knowledge capture as "a process by which the expert's thoughts and experiences are captured". Milton expands on this definition arguing that there are two stages to knowledge capture - *eliciting knowledge* which is the capture of knowledge that is not already in the knowledge base and *validating knowledge* which involves checking knowledge in the knowledge base is correct, complete and relevant.

There are three main steps involved in knowledge capture:

- Using an appropriate tool to elicit the information from the expert
- Interpreting the information and inferring the expert's underlying knowledge and reasoning process
- Using the interpretation to build the rules that represent the expert's thought processes or solutions

(Awad and Ghaziri 2004)

While knowledge sharing "the process of exchanging knowledge (skills, experience and understanding) among researchers, policymakers and service providers" (Tsui 2006, p.5). When combined these separate definitions of knowledge management components create the overall concept of knowledge management.

Introduction to Communities of Practice

A community of practice is defined as:

"Groups of people who share a concern, a set of problems, or a passion about a topic, and who deepen their understanding and knowledge of this area by interacting on an ongoing basis" (Wenger et al. 2002, p.4). Peltonen & Tuija (2004) argue that there are three components to this definition which defines a community of practice within three dimensions:

What is it about - it's a *joint enterprise* as understood and continually renegotiated by its members.

How it functions) - *mutual engagement* that bind members together in a social entity.

What capability it has produced – the *shared repertoire* of communal resources (routines, sensibilities, artefacts, vocabulary, styles etc) that members have developed over time.

Nichani (2001) proposes three types of groups that members of communities practice are a part of:

Connectors who know everyone and are very good at making friends, have very good social skills and self-confidence and are always curious about what is going on (Nichani 2001; Saint-Onge and Wallace 2003).

Mavens who connect different people with information and collect information that might be useful for others (Saint-Onge and Wallace 2003).

Salesmen are the persuaders, they take the responsibility of persuading other members of the community into accepting something new or changing their minds, they are good at expressing emotions (Nichani 2001; Saint-Onge and Wallace 2003).

Every member of the community of practice is grouped into one of the three groups above based on their role.

Knowledge is what forms a community of practice and new knowledge is created based on the practices experience and shared and existing knowledge is shared and decisions are made by the community based on this knowledge. Saint-Onge & Wallace (2003) argue that when taking this into consideration the community of practice is then based on three components: access to existing knowledge, knowledge exchange/transfer and the creation of knew knowledge.

Access to existing knowledge that is primarily codified or explicit (eg. knowledge objects stored in a database) (Saint-Onge and Wallace 2003).

Knowledge exchange gained through sharing experience that is primarily tacit, but may also be explicit; a validation of information (eg. conversations in the community) (Saint-Onge and Wallace 2003).

Creation of new knowledge through collaborating on innovations (eg. result of a problem-solving exercise based on a productive inquiry) (Saint-Onge and Wallace 2003).

Knowledge Management and Communities of Practice – How does it all fit?

When members of a community of practice get together it is a time to share the knowledge they have generated and collected, improve current processes within the organisation, find solutions to problems, test theories and discuss concerns, challenges or issues. It allows members to test ideas on other COP members and create an understanding amongst community members. Knowledge gained by community members is then used and dispersed among those in the practice.

Communities of practice have not yet been applied to AFL clubs nor have guidelines for knowledge sharing and transfer as shown by a gap in the literature.

3. METHODOLOGY

An AFL club has been used as a case study and is the basis of this research. In depth interviews have taken place with members of the football department and support staff including the: Senior coach, assistant coaches, line coaches, fitness staff, medical officers, trainers, information technology and video analysts. Participants were asked detailed questions about the clubs business processes and what their role in these is, the types of data they share and generate, whom they communicate with and how they go about it as well as how often and about what.

Interview data was 'coded' by category and theme allowing the researcher to delve deeper into the data and aggregate results based on category or theme for example knowledge transfer. This was done using Atlas.ti to identify the important components and relationships between staff members and processes as well as confirming previously established theories about the data and it's relationships. Post-interviews business process diagrams were developed to show the process flow of information, communication and data storage throughout the organisation as well as establish the goal and outcomes of each unique process. Communities of practice (COPs) were then identified from staff that were involved in each unique process and their primary area of involvement as well as the aim of the overall process.

4. DISCUSSION

Using Business Process Modelling Notation (BPMN 2.0) business process maps (BPM) have been generated for each individual process the members of the football department are involved in. Examples include - Providing live feedback to players; eligibility of injured players for selection; success or failure of KPI's during games; stoppage analysis; midfield review of opposition teams and off-field player development. An example of the 'player feedback provided by line coaches and the identification of welfare issues' can be seen in Figure 1.

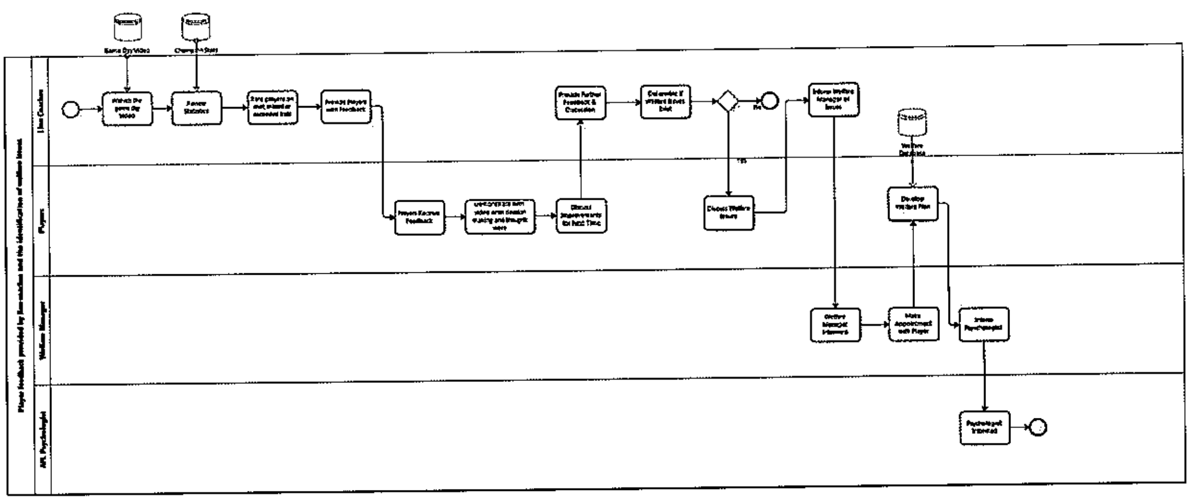


Figure 1: An example of the 'player feedback provided by line coaches and the identification of welfare issues

The process begins by line coaches (forward, midfield, ruck, defence) watching game day video which is accessed from a data store as well as

reviewing the statistics provided to all the AFL clubs from Champion Data which is similarly stored in a data store.

Line coaches then proceed to rate each individual player in their line that participated in the game on 'player traits' these are 'traits' or 'components' that have been decided on by coaching staff earlier in the year as well as the inclusion of players in selected traits. An example of this could be that forward line player A is expected to maintain a 70% successful goal rate or lead for the ball into a certain position on the field 20 times during a quarter. Coaches rate each player on whether they have met, missed or exceeded the individualised traits that have been set for them to achieve during the game. This feedback is then provided to players in their individual line meetings by coaching staff, video examples are shown and a player will describe and discuss their decision making throughout the clip and how they felt about the situation as well as what they can do to improve next time they are in the same situation. Coaching staff then discuss the decision making process with player and provide further feedback.

While coaching staff are providing feedback and having informal discussions with a player are taking place line coaches role is also to identify if the player is having any welfare issues, an example of this could be problems at home with their partner or difficulty with current living arrangements. If the player is not experiencing any welfare issues the meeting concludes once feedback and discussion has concluded. If the player is experiencing welfare

issues once the line coach has identified the issue it is their responsibility to then inform the player welfare manager. An appointment is then made with the player by the welfare manager in which a welfare plan is developed to assist with resolving the welfare issue in question, this plan is stored in the welfare data store for future reference. The final step of the process involve the notification of the AFL psychologist allocated to the club.

As seen in Figure 1 there are three groups of people involved (excluding the players); the line coaches, player welfare manager and AFL psychologist. As per Wenger et al (2002, p.4) a community of practice is "Groups of people who share a concern, a set of problems, or a passion about a topic, and who deepen their understanding and knowledge of this area by interacting on an ongoing basis"

The development of these models and the role that football department staff play in each individual business process has led to the discovery of a number of communities of practice within the department football department. This can be seen in Figure 1 in which the personal involvement of the line coach, player welfare manager and AFL psychologist and the areas of involvement, tactical

and strategic analysis and player welfare combine to create a community of practice.

The two areas/groups share a concern (player welfare) and share and deepen their knowledge in this area by coming together interacting and discussing the issues on a regular basis as per Wenger et al (2002) definition previously discussed. Similarly with an example of the 'player receives injury during game' set of business processes you have personal involvement from doctors, the player, the senior coach and the line coach in the areas of injury management, tactical and strategic analysis and player rehabilitation. These areas also come together to form their own community of practice which is interested in looking after the injured player and aiding in their return to the field during the game if possible and if not their rehabilitation and injury management sharing information about how this will occur and the best methods on a regular basis.

The football department is made up of many of these communities of practice in which all share a concern in various areas of the football club and who come together to interact and share their knowledge with the rest of the community. The presence of these communities of practice within the football department means that many different groups of people who look after or are 'experts' in different areas can come together with people whose goal is the same to share and transfer knowledge and contribute to an area of concern or interest rather than simply focusing on their specific area such as

psychology they can become involved in the overall picture and problem on a larger scale.
It has often been believed that communities of practice within an organisation increase communication and knowledge and workflow but due to the abstract nature of communities of practice their impact and is challenging to quantify. This can be seen below in which Stolovitch & Keeps (2006) categories the benefits of communities of practice within three areas: for business, for community and for the individual.

Communities of practice not only help drive business strategy but help problem solving both within local divisions and organisation wide. It assists in building knowledge bases throughout the organisation and increases the opportunity for knowledge sharing and innovation whilst assisting employees with their jobs. Whilst providing a sense of community to individuals it assists in developing their communication and problem solving skills in a nurturing environment allowing them to stay up to date with current information and apply existing knowledge to problems or situations that are being examined by the community of practice.

As seen above these benefits are intangible and cannot be easily quantified by the organisation or individuals as to their impact. Wenger, McDermott & Snyder (2002) argue that it is important to support communities of practice and cultivate the efforts employees are making to share and transfer knowledge and resources throughout the organisation rather than killing the initiative with procedural red tape and departmental constraints. Rather than restricting employees and their enthusiasm they need to channel it into a constructive process such as a community of practice which will not only benefit the individual but the overall organisation in turn making those involved feel valued.

As it can be seen through the demonstrated benefits of communities of practice if implemented within an AFL football department the example provided earlier in the paper would not occur again as the fitness staff would be attending the same meetings as the doctors, line coaches, senior coach and physios which would prompt knowledge sharing regarding the GPS data being below average for the player. The fitness staff would speak to experts in the area with the same common goal in order to determine what the issue with the player is, for example - they are developing an injury but it has not fully occurred yet rather than a general assumption that the player is not trying or their game isn't peaking at the correct time. Rather than having information passed on second or third hand staff can communicate face to face about any issues that are occurring or questions they might have regarding players as well as well as sharing this wealth of information amongst each other it will assist each individual to not only provide the players with the best service possible but to ensure they are in full health and are competitive on the field. Information being stored correctly and being provided to those who require it within a timely manner is critical in any organisation especially a football department where the majority of information is time sensitive.

5. CONCLUSIONS

Communities of practice are essential in business for sharing and transferring knowledge throughout the organisation and encouraging shared problem solving and a sense of value and involvement. The fostering, support and generation of communities of practice ensures that the necessary people receive the information they need in a time critical manner and allow communication throughout the community of practice to resolve the issue or problem that is occurring rather than one expert in one area make a decision, communities of practice gather all the experts in the relevant areas that can assist with the resolution of the problem. Therefore it is recommended that AFL clubs foster and implement communities of practice to assist with the timely sharing and transfer of knowledge amongst department members and resolution of problems.

Acknowledgements

I wish to thank the AFL club who participated in my PhD case study research.

References

Awad, E., and Ghaziri, H. 2004. *Knowledge Management*, (1st ed.). New Delhi: Pearson Education.

Collins, H. 2010. *Tacit and Explicit Knowledge*, (1st ed.). Chicago: The University of Chicago Press.

Magalhaes, R. 2004. *Organizational Knowledge and Techology - an Action-Oriented Perspective on Organization and Information Systems*, (1st ed.). Massachusetts: Edward Elgar Publishing Limited.

Mair, R. 2004. *Knowledge Management Systems - Information and Communication Technologies for Knowledge Management*, (2nd ed.). New York: Springer-Verlag.

Nichani, M. 2001. "Knowledge Management Core." Retrieved 4 July, 2012, from http://www.elearningpost.com/elthemes/kmcore.asp

Peltonen, T., and Tuija, L. 2004. "Communities of Practice and the Social Process of Creation: Towards a New Vocabulary for Making Sense of Organizational Learning," *Problems and Perspectives in Management*:4), pp 249-262.

Saint-Onge, H., and Wallace, D. 2003. *Leveraging Communities of Practice for Strategic Advantage*, (1st ed.). Massachusetts: Elsvier.

Sallis, E., and Jones, G. 2002. *Knowledge Management in Education - Enhancing Learning and Education*, (1st ed.). New York: Routledge.

Scarborough, H., Swan, J., and Preston, J. 1999. *Knowledge Management: A Literature Review, Issues in People Management*, (1st ed.). London: Institute of Personnel and Development.

Stolovitch, H., and Keeps, E. 2006. *Handbook of Human Performance Technology - Principles, Practices and Potential*, (3rd ed.). San Francisco: Wiley.

Tsui, L. 2006. *A Handbook of Knowledge Sharing: Strategies and Recommendations for Researchers, Policymakers and Service Providers*, (1st ed.). Alberta: Community University Partnership for the Study of Children, Youth and Families.

Von Krough, G., Nonaka, I., and Nishigushi, T. 2000. *Knowledge Creation: A Source of Value*, (1st ed.). New York: Palrave.

Wenger, E. 2006. "Communities of Practice: A Brief Introduction." Retrieved 30 May, 2012, from http://www.ewenger.com/theory/

Wenger, E., McDermott, R., and Snyder, W. 2002. *Cultivating Communities of Practice: A Guide to Managing Knowledge*, (1st ed.). Boston: Harvard Business School Press.

Wenger, E., and Snyder, W. 2000. "Communities of Practice - the Organizational Frontier," in: *Harvard Business Review*. Massachusetts: Harvard Business School Publishing, pp. 139-145.

"Some people think football is a matter of life and death... I can assure them it is much more serious than that" ~ ***Bill Shankly.***

"They think it's all over; it is now"~ ***Kenneth Wostlenholme,*** before and after England went 4-2 up against West Germany in 1966.

"People ask me what I do in winter when there's no baseball. I'll tell you what I do. I stare out the window and wait for spring." ~ ***Rogers Hornsby***

"Doctors and scientists said breaking the four-minute mile was impossible, that one would die in the attempt. Thus, when I got up from the track after collapsing at the finish line, I figured I was dead." ~ ***Roger Bannister.***

"I've missed more than 9,000 shots in my career. I've lost almost 300 games. Twenty-six times I've been trusted to take the game-winning shot and missed. I've failed over and over and over again in my life. And that is why I succeed" ~ ***Michael Jordan.***

"Pressure? I'll tell you what pressure is. Pressure is a Messerschmitt up your ****. Playing cricket is not" ~ ***Keith Miller.***

"Hockey is a sport for white men. Basketball is a sport for black men. Golf is a sport for white men dressed like black pimps" ~ ***Tiger Woods.***

"There is water in every lane, so it is OK" ~ swimmer ***Ian Thorpe*** on being drawn in lane five.

"Whoever said 'It's not whether you win or lose that counts' probably lost" ~ ***Martina Navratilova.***

"You can make a lot of money in this game. Just ask my ex-wives. Both of them are so rich that neither of their husbands work" ~ ***Lee Trevino.***

"The bowlers Holding the batsman's Willey" ~ a true but hilarious piece of commentary from the BBC's ***Brian Johnston.***

"When considering the stature of an athlete or for that matter any person, I set great store in certain qualities which I believe to be essential in addition to skill. They are that the person conducts his or her life with dignity, with integrity, courage, and perhaps most of all, with modesty. These virtues are totally compatible with pride, ambition, and competitiveness" ~ ***Don Bradman.***

"The fight is won or lost far away from witnesses – behind the lines, in the gym, and out there on the road, long before I dance under those lights" ~ One of many from the eternal boxing legend ***Muhammad Ali.***

"It's not the size of the dog in the fight, but the size of the fight in the dog ~ ***Archie Griffen.***

"The difference between the old ballplayer and the new ballplayer is the jersey. The old ballplayer cared about the name on the front. The new ballplayer cares about the name on the back ~ ***Steve Garvey.***

"That's great, tell him he's Pele and get him back on" ~ ***John Lambie,*** when told his concussed striker could not remember who he was.

"We had a good team on paper. Unfortunately, the game was played on grass" ~ the indomitable ***Brian Clough.***

"I think my favourite sport in the Olympics is the one in which you make your way through the snow, you stop, you shoot a gun, and then you continue on. In most of the world, it is known as the biathlon, except in New York City, where it is known as winter" ~ ***Michael Ventre.***

"The trouble with referees is that they just don't care which side wins" ~***Tom Canterbury.***

"Awards become corroded, friends gather no dust" ~ ***Jesse Owens.***

APPLICATIONS OF MATHEMATICS AND COMPUTERS IN EXERCISE AND SPORT SCIENCE: HISTORICAL PERSPECTIVE THE AUSTRALIAN EXPERIENCE

Ian Heazlewood[a]

[a] *Charles Darwin University*
[a] *Corresponding author: ian.heazlewood@cdu.edu.au*

Abstract

Mathematics and computers applications have evolved over the past 25 years in exercise and sport science. Bachelor's degrees were gradually introduced over the past 25 years and the dependence on mathematics and computers has evolved from a novelty to high dependence. For example, from simple software programs to computers to transform with analogue to digital signal analysis via force platforms, metabolic carts to assess O_2 and CO_2, respiration rates and volumes and ECG; software for motion analysis and mathematical solutions related to linear and angular kinematics to today, where they are the driving platforms for E-learning, computer assisted learning, instrumentation and statistical and signal analysis via multivariate software packages, signal analysis and signal transformations. The question is with the current high dependence on computers to derive mathematical and statistical solutions, as well as now an indispensible component in delivery of university courses within Australia, have we actually improved the skill set of exercise and sports scientists and training related to applications of mathematics and computers to solving and understanding exercise and sport science research questions?

Keywords: mathematics, computers, history, exercise, sport science Australia

1. INTRODUCTION

Mathematics and computers are developing a central focus in exercise and sports science, from an initial novelty item in the mid-1980's in laboratories and providing limited computer assisted learning to today, where in 2012 mathematics and computers are indispensible tools for exercise and sport students and scientists. In higher education in the exercise and sport sciences it is now a curriculum expectation that students will have knowledge and skills in mathematics and computers.

Applications of mathematics and computers in exercise and sport science have varied over the years from arithmetic, mathematical, physics, multivariate statistical and neural networks applications.

1980's

In the mid 1980's applications were in the context of undergraduate textbooks and research were related to kinesiology, physical education and exercise was identified in 1987 (Baumgartner & Jackson, 1987) :

- Arithmetic in exercises physiology using basic arithmetic calculations (Johnson & Nelson, 1986).
- Mathematical calculations in biomechanics based on physics and applied mathematics principles, such as kinematics (displacement, velocity, acceleration), dynamics (forces, torques), kinetics (energy), projectile motion (jumping, throwing events), fluid dynamics (fluid drag, buoyancy, swimming, rowing), statics (centres of mass, centres of pressure) (Luttgens & Wells, 1982; Hay, 1985; Hay & Reid, 1982). Some derivations based on first principles differentiation and integration.
- Statistical applications basic descriptive statistics (mode, median, means, standard deviation, range, variance) and inferential univariate applications (chi square, t–tests, one-way

ANOVA between groups, correlations, both Pearson and Spearman, and bivariate regression) (Johnson & Nelson, 1986; Baumgartner & Jackson, 1987; Safrit & Wood, 1986). Problem solving calculations based on hand held calculators. Students and staff had to know the mathematical/arithmetic calculation steps.

- Some software development in statistical problem solving such as staff using Applesoft Basic to develop and write some simple statistical programs (Rothstein, 1985).
- Some mainframe access was available to conduct more complex mathematical solutions based on multivariate statistics and computations such as early versions of SPSS (Rothstein, 1985; Norusis, 1985).
- Instrumentation based on hardware, primary variable display and no computer interface. Secondary variable calculations required to derive biomechanical (velocity, acceleration) and exercise physiological constructs (oxygen pulse, VO_2 max.).

1980's Teaching and Learning

Some dependence on arithmetic, mathematics and computers in class in terms of face to face or internal teaching, using computers to facilitate learning or computers used as a resource/instrument to promote learning. Initially computers were a novelty in the exercise and sport science classroom and mathematics and statistics in exercise and sport science perceived by students as a challenge. Basically, 1980's consisted of some computer assisted learning as simple games and some simple interactive statistical programs based input-output approach with undergraduates. Such e-learning-computer driven assisted learning is the exception as face to face teaching is the norm.

1990's

The 1990's were evolving in terms of PC power and as a consequence instrumentation or hardware was coupled with computer interface to collect, compile, transform and display data. Secondary variables derived were based on software capability provided by manufacture (Morgan Gas analysis system, motion analysis systems and CYBEX muscle evaluation systems). In terms of discipline content the exercise and sport science content in the university sector essentially the same in terms of content. Textbook authors writing 3rd, 4th and 5th editions (Hay, 1993; Luttgens & Hamilton, 1997)) or new versions of previous ideas (Bloomfield, et al., 1992; Bloomfield et al., 1994; Keighbaum & Barthels, 1996).

Some dependence on arithmetic, mathematics and computers in class in terms of face to face or internal teaching, using computers to facilitate learning or computers used as a resource/instrument to promote learning. Initially computers were a novelty in the exercise and sport science classroom and mathematics and statistics in exercise and sport science perceived by students as a challenge. Basically, 1980's consisted of some computer assisted learning as simple games and some simple interactive statistical programs based input-output approach with undergraduates. Such e-learning-computer driven assisted learning is the exception as face to face teaching is the norm.

Paper based systems used in distance-external education delivery.

2000-2010 High Dependence on Mathematics, Computer Hardware and Software

Software

More sophisticated statistical and mathematical software developments for example, SPSS, MATLAB, statistica, and neural network programs. Increased functionality and calculation speeds enabled by PC 'evolution' and as CPU's become more powerful, as well as significant increases in RAM space.

"**MATLAB** (**mat**rix **lab**oratory) is a numerical computing environment and fourth-generation programming language. Developed by MathWorks, MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, Java, and Fortran" and frequently used in exercise and sport science in signal analysis (http://en.wikipedia.org/wiki/Matlab; http://www.mathworks.com.au/academia/student_version/).

Unfortunately, most of the mathematics is black boxing where student and researcher input syntax and functions via drop down boxes, define experimental parameters (drop down boxes) such as sampling rates, setting graphical axes in terms of display output and selecting method of presentation of results (2-D and 3-D rotations).

In the different disciplines:

- Arithmetic in exercises physiology using basic arithmetic calculations (ACSM, 2012) mostly completed by software manipulating primary variables and other calculations are conducted by software that interfaces with the hardware or raw data (non-transformed data), such as MOXUS O_2, CO_2 and ventilation system. Such as calculation of oxygen pulse and respiratory exchange ratios. (RER).

- Mathematical calculations in biomechanics based on physics and applied mathematics principles, such as kinematics (displacement, velocity, acceleration), dynamics (forces, torques), kinetics (energy), projectile motion (jumping, throwing events), fluid dynamics (fluid drag, buoyancy, swimming, rowing), statics (centres of mass, centres of pressure) (references) calculations (references) mostly completed by software manipulating primary variables linear and angular displacement with time velocity and acceleration and two and three dimensional transformations (2-D and 3-D space) software derived (VICON, SIMI motion analysis systems).
- Multivariate statistical analysis heavily dependent upon statistical software, such as SPSS multivariate, Statistica (multivariate) AMOS (structural equation modelling), model defining software where mathematical steps solutions derived by software which is fast, convenient but black boxing for most end-users. Honours (multivariate) course work to9 teach these multivariate skills and postgraduate students learn applications to solve specific research problems (non course work based).

1990's Teaching and Learning

Development of some e-learning and computer driven computer assisted learning in class. To cater for distance/external delivery of university subjects and degrees predominantly paper based.

2000-2010

No revolution as in no significant changes in the disciplines in terms of applications of arithmetic, mathematics, multivariate statistics and computers in the disciplines concerning discipline based problems in exercise and sports science, such as sports medicine, exercise physiology, biomechanics and motor learning (Brooks, et al., 2005; American College of Sports Medicine, 2006; 2010). In these disciplines it is the hard measurement with soft statistics model of problem solving. Textbooks essentially use the mathematics of the 1980' and 1990' as mentioned previously. Sport psychology is becoming more multivariate in nature, the soft measurement and hard statistical analysis model. Some slight improvements in computer driven instrumentation in terms of data acquisition across disciplines and lots of black boxing in this context.

Teaching and Learning

The revolution is upon us. Computer assisted learning using sophisticated e-learning platforms.

Current model driven by e-learning platforms that are computer driven-dependent following significant developments of commercially available e-learning platforms, such as WEB-CT, Blackboard and Moodle.

"**Moodle** (abbreviation for Modular Object-Oriented Dynamic Learning Environment) is a free source e-learning software platform, also known as a Course Management System, Learning Management System, or Virtual Learning Environment (VLE). As of December 2011 it had a user base of 72,177 registered and verified sites, serving 57,112,669 users in 5.8 million courses," (Wikipedia, 2012, http://en.wikipedia.org/wiki/Moodle).

Moodle was developed to assist and enhance educators to create online courses with a focus on interaction and collaborative construction of content and the processes and software are in continual evolution. The first version of Moodle was released over 10 years ago in 2002. It all seems so new.

Today some Australian universities have 70% plus students undertaking undergraduate and postgraduate coursework degrees in distance or external mode. For example, the Bachelor of Exercise and Sport Science degree at Charles Darwin University is delivered in both internal (face to face) and external (distance) mode. The first course in Exercise and Sport Science to be delivered completely in this distance-external mode in 2012.

In this context, "the education revolution", LearnLine or e-learning support is where face to face teaching is supported by computer assisted e-learning.

LearnLine reliant is where there is no face to face teaching and students are in distance/external education mode is where all interactions are by computer assisted e-learning.

Delivery of lecturing content such as PowerPoint, uploaded videos, pdf files, MP3 AV files, links to various websites, uploading downloadable print materials, emails, discussion boards, upload assessments, deliver an assessment, track assessments, track student interactions, return assessments, conduct interactive tests, set release and closure dates re resources and so on. Interact with students in synchronous and asynchronous modes.

- Applications in undergraduate statistics subjects/units using more sophisticated statistical software, student access to latest versions SPSS. Student-staff-computer interactive learning environment.

- Significant support computer and staff investment, such as on site student computer labs, library computer labs, staff functionality (all academic staff). Expectations that all students have functional computers that can interact with university LearnLine e-learning platform.
- Usual access to twitter and face book and emphasis and now intensive LearnLine professional development, drop-in LearnLine sessions, extensive LearnLine support staff to assist staff and students (technical support), LearnLine site audits re compliance of staff sites prior to student release, quality audits of sites by learning and teaching sections such as Deans having access to all staff LearnLine sites, ranking of LearnLine sites re "evidence of best practice", use of LearnLine system to micromanage staff. Delivery and collecting of data via student SELTS (student evaluation of learning and teaching of staff delivering units/subjects), delivery of
- SELTS scores to staff and heads of school for review, a data dump.

2012 – Where Are We Now?
Exercise and Sport Scientist at Work

- Get to work turn on computer, "talks" to server, sign in password approved, access granted.
- Open emails, Review emails, Conduct research review websites, databases via virtual library.
- Plug in one terabyte portable hard drive with all files.
- If with the special interest groups using mathematics and computers in sport then open and analyse research database, input data if required. Run statistical solution open software SPSS, statistica, neural networks, matlab, AMOS as required to solve appropriate research design question.
- Write report, open wordfile, save file and close.
- Review budget excel spreadsheet, tweak budget, save and close.
- Log onto staff LearnLine site. Open this week's learning materials, enable access for students via date release, or set time release 3 months previously. Go to grade centre to assess ho has submitted assignments (softcopy submissions no hardcopy submissions), note time and date of submission (late assignment?). Download assignments to another file to mark. Mark softcopy, return marked assignment to students via softcopy pdf file via LearnLine. Receive additional feedback from students via LearnLine emails.
- Add additional learning materials if required, announcements, posts, and so on. Internal students (face to face) see in lectures and practicals. See external students in person very occasionally if at all as communication now via computer interface.

Today a high dependence on e-learning results in high dependence on functional computing so exercise and sports scientists now have to undergo professional development with e-learning and compulsory pedagogical (children teaching and learning theory and skills) and androgogical (adult teaching and learning theory and skills) training.
Significant focus on up skilling staff re teaching and learning such as professional development in:

- Learning Teaching Induction at CDU – e-learning
- Blackboard Collaborate– e-learning
- Building Group work Skills – e-learning
- Introduction to Learnline: Assisting students– e-learning
- Introduction to Learnline: Creating units– e-learning
- Online Teaching with Learnline– e-learning

Graduate Certificates in Teaching and Learning in Higher education now require e-learning competence such as; Overview-A brief introduction to online leaning and Learnline, from both the teacher's and the student's perspective; Who should attend-Required as part of probation for all new Academic level A and B staff; and Delivery-Flexible online delivery. Such e-learning-computer driven assisted learning is now the norm not the exception.

Roles in Professional Associations-Learned Societies Emphasising Mathematics and Computers in Exercise and Sport Science

Learned societies such as Institute of Mathematics and Its Applications (IMA), International Association of Computer Science in Sport (IACSS) and MathSport frequently hold international and national conferences that focus on the application of mathematics in exercise and sports science. "MathSport is a special interest group of ANZIAM. The group consists of a loose forum for Australian and New Zealand sports scientists to interact. MathSport holds biennial meetings; the Mathematics and Computers in Sport Conferences," (http://www.anziam.org.au/Mathsport).

However, an important observation is the majority of members and conference delegates of these learned societies are not from an exercise and sport science background and most are from mathematical, IT, engineering, statistical backgrounds. The later situation is a good fit based on member and delegate training, but the former situation of low numbers of exercise and sports scientists trained in concerning. Why? Exercise and sports scientist believe that exercise and sport science research question are addressed by exercise and sport scientists, yet training in the more complex quantitative aspects of exercise and sports science are not usually taught within these programs delivered by Australian universities as revealed by web searches of Australian university subjects that make up the content of exercise and sport science courses.

As a consequence a wealth of information, enabling more substantive quantitative insights into complex human movement phenomena is not addressed. In this context the valuable information is hidden within the data sets. So what do we do as members of these special interest groups within learned societies?

1. Promote these more insightful mathematical and computer based problem solving approaches within the exercise and sports science domain.

2. Communicate more assertively these research possibilities and outcomes to our colleagues in the exercise and sport science disciplines.

3. Encourage the development of subjects based on mathematical and computer based problem solving within Australian delivered exercise and sport science degrees.

4. Present such approaches as very exciting and as a new way of looking and evaluating the real complexity within human exercise and sport science.

5. Expand our memberships by actively encouraging exercise and sport scientist trained professionals to join our team.

6. Being more assertive at multidisciplinary exercise and sports science and sports medicine international and national conferences with specific emphasis discipline streams/sessions as well as "mixing it up" with the other disciplines to indicate how we contribute exercise and sport science research solutions.

5. CONCLUSIONS

The developments in terms of applications of mathematics and computers in exercise and sport science teaching content have not changed significantly in terms of actual mathematical knowledge expected or required. Today, the reality is the mathematical algorithms are usefully selected for the exercise scientist within the black box software in terms of data acquisition. More complex mathematical problem solving is not taught within many Australian degrees based on exercise and sports science, in some sense this approach is a novelty for exercise and sport scientist or it is not known.

Computers and computer software are driving the show via manufacturers who provide instruments in the market of exercise and sport science.

In terms of teaching and learning a revolution has occurred in terms of delivery via e-learning platforms and computer functionality and whether this results in improved research problem solving and up-skilling exercise and sports scientist's remains to be confirmed. In terms of actual mathematical applications in the exercise and sport science context and insights based on university curriculum development and available texts in sports exercise and science conceptual developments and skill outcomes have not really evolved.

References

American College of Sports Medicine – ACSM. (2006). ACSM's Guidelines for Exercise Testing and Prescription. (7th Ed.) Sydney: Wolters Kluwer.

American College of Sports Medicine – ACSM. (2010). ACSM's Health-Related Physical Fitness Assessment Manual. (3rd Ed.). Sydney: Wolters Kluwer.

Baumgartner, T. & Jackson. (1987). Measurement for Evaluation in Physical Education and Exercise Science. Dubuque, IA; Wm. C. Brown.

Bloomfield, J., Fricker, P. & Fitch, K. (1992). Textbook of Science and Medicine in Sport. Melbourne: Blackwell Scientific Publications.

Bloomfield, J., Ackland, T. & Elliot, B. (1994). Applied Anatomy and Biomechanics in Sport. Melbourne: Blackwell Scientific Publications.

Brooks, G., Fahey, T. & Baldwin, K. (2005). Exercise Physiology: Human Bioenergetics and its Applications. (4th Ed.). Sydney: McGraw – Hill.

Hay, J. (1985). The Biomechanics of Sports Techniques. (2nd Ed.). Englewood Cliffs, N.J.: Prentice-Hall.

Hay, J. (1993). The Biomechanics of Sports Techniques. (4th Ed.) Englewood Cliffs, N.J.: Prentice-Hall.

Hay, J. & Reid, J. (1982). The Anatomical and Mechanical Bases of Human Motion. Englewood cliffs, N.J.: Prentice-Hall.

Johnson, B. & Nelson, J. (1986). Practical Measurements for evaluation in physical Education. Edina, MN: Burgess Publishing.

Keighbaum, E. & Barthels, K. (1996). Biomechanics: A Qualitative Approach for Studying Human Movement. (4th Ed.). Sydney: Allyn & Bacon.

Luttgens, K. & Wells, (1982). Kinesiology: Scientific Basis of Human Movement. New York: CBS College Publishing.

Luttgens, K. & Hamilton, N. (1997). Kinesiology: Scientific Basis of Human Movement. (9th Ed.). Dubuque, IA: WCB/McGraw-Hill: New York: CBS College Publishing.

McLester, J. & St. Pierre, P. (2008). Allied Biomechanics: Concepts and Connections. Australia: Thompson-Wadsworth.

Norusis, M. (1985). Advanced Statistics Guide: SPSSX. Sydney: McGraw-Hill.

Rothstein, A. (1985). Research Design and Statistics for Physical Education. Englewood cliffs, N.J.: Prentice Hall

Safrit, M. & Wood, T. (1986). Introduction to Measurement in Physical Education and Exercise Science.

Schell, J. & Leelarthaepin, B. (1994). Physical Fitness Assessment in Exercise and Sport Science. (2nd Ed.). Sydney: Leelar Biomediscience Services.

SAFETY INDICES FOR SURF LIFE SAVING COMPETITIONS

Neville de Mestre[a, c], *Gary McCoy*[b]

[a] *Emeritus Professor of Mathematics, Bond University*
[b] *Senior Teaching Fellow*
[c] *Corresponding author:Neville_de_mestre@staff.bond.edu.au*

Abstract

Fatalities at the Australian Surf life Saving Championships in recent years have prompted us to develop a Competition Surf Safety Index for each category of competitor at any surf carnival held on any beach at any time. The model for the index is based on the difference between the skills' rating of a competitor and the hazard rating of the competition surf zone. When the category's index is negative, it indicates that conditions are too dangerous for that group of competitors.

---

1. INTRODUCTION

There have now been three male fatalities at the Australian Surf Life Saving Championships in the past sixteen years, two of them occurring as recently as 2010 and 2012. It therefore seems appropriate to develop some quantitative Surf Safety Indices for the water events at any beach where a surf carnival is to take place. Such indices may vary with the different levels of experience of the relevant competitors because of their age, gender and skill.

It should be noted at this point that, by virtue of their training and fitness level, the skill level of surf life saving competitors is much higher than that of the general public. Short and Hogan (2006) suggested a beach hazard rating system with "...61% of (Australian) beaches rating 5 or higher on a 1 to 10 scale...". Although relevant for the general public, this is not directly applicable to competitors at a surf carnival venue.

Surf carnivals have been held for the last one hundred years, usually in small and moderate surfs, but sometimes in large and relatively dangerous surf. A history of the Australian Surf Championships (Galton, 1984) reveals that "mountainous" seas occurred in 1915, 1946, 1947, 1948, 1955, 1959, 1963, 1964, 1969, 1970 and 1977. No lives were lost on any of these occasions. Since then there have been big seas in 1996, 2000, 2005, 2010 and 2012. For some of these the carnival events were shifted to a safer beach. Clearly a more robust and helpful quantitative measure of risk is needed to assist organisers in reducing the potential of an unfortunate accident happening to a competitor in any of these events.

The ideas in this paper are based on the authors' academic and surf life saving expertise. We are internationally recognised applied mathematicians in the fields of fluid mechanics and statistics. Together, we have over 100 years of surf life saving experience, involving patrols, rescues and surf awareness education. We have both competed, and been successful, at State and Australian Open and Masters' level of competition, including all current forms of swimming and craft events (except for boats), some in conditions now classified as mountainous and/or dangerous.

2. THE MODEL

By taking into account the higher skill level of surf life saving competitors, a Competition Surf Safety Index (CSSI) will be developed for the combined age, gender and skills category of each proficient competitor. It will consist of a Competitor Surf Experience Rating (CSER) from which is subtracted a Surf Hazard Rating (SHR). Thus

$$CSSI = CSER - SHR \qquad (1)$$

When the CSSI produces a negative value for a category, it can be used as an indicator that either,

(i) those competitors and their cohorts should be moved to a safer beach, or

(ii) competition for them should be suspended until conditions become less dangerous.

As a corollary, for each competition category where the CSSI ≥ 0, the risk of an accident happening in those water conditions during competitions is as low as is reasonably practicable.

3. SURF HAZARD RATING (SHR)

For surf life saving competitors, there are hazards due to sharks, stingers, submerged rocks and the movement of the water itself. As ground swell in the form of water wave trains approaches a coastline, it starts to interact with the gently sloping ocean bed; each wave in the train loses its symmetrical profile and starts to steepen on the front face. Eventually the wave breaks forming a recognisable surf zone of rolling hydraulic bores or surf fronts, which can propel humans and craft towards the shoreline.

To develop a quantifiable model of the hazards due to water movement, the main characteristics of the potentially hazardous surf zone are identified as:

(i) the wave height, type and period

(ii) the surf zone length

(iii) the turbulence of the water surface

(iv) the tide

(v) the orientation of the ground swell to the local coastline

(vi) any directional movement of the water body.

The latter will include outward flows (rips) and sideways drags (littoral drifts). These characteristics may change over a number of hours, but it is their possible maximum effect on surf competitors' safety that should be incorporated into any Surf Hazard Rating calculation within a particular hour.

The cause of the most dangerous surf zone characteristics is usually storms at sea which generate the moving wave trains (ground swell) that eventually impinge on the coastline. The ground swell is frequently amplified by the prevailing winds that accompany the weather patterns. Extreme ground swell can be due to severe low depressions in atmospheric pressure many kilometres away from the coastline. Sometimes these produce cyclonic conditions and huge seas along a coastal strip. Local storms may also contribute to the surf zone characteristics on any beach, and produce major or minor changes depending on the geomorphology of the beach, its orientation to the wind direction, and the interaction of the ground swell with the wind direction. On rare occasions an underwater earthquake can also produce a significant ground swell towards the coastline.

Therefore, although the wind produces the waves in general, and the coastline produces the surf zone through the decreasing depth of the water as the waves approach the shore, our surf zone characteristics will not include the wind explicitly. The effect of the wind will be included indirectly in the measurements of wave height, wave period, surf zone length, directional movement of the water, and the interactive surface turbulence. From our experience it is clear that surf conditions can change rapidly overnight, and often from hour to hour with the change in tide or the arrival of a weather front. Consequently the SHR may be changing quite frequently at any beach over the passage of time. In Tables 1 to 7 listed below, a quantitative danger value is assigned to each measurement class of the major surf zone characteristics. The rating values are all positive integers and begin with zero for no contribution to danger for the competitors. An effort has been made to ensure that the rating values across the five tables are relatively equivalent and therefore consistent.

Further tables may be added to the SHR section of the model at a later date based on further research together with an analysis of the Australian Beach Safety and Management Program database (ABSAMP, 1990), but it seems that this initial set of tables caters for the majority of surf zone changes that influence safety for surf life saving competitors.

3.1 Wave Height Rating (WHR)

The accepted definition of wave height when discussing surf dynamics is the difference between the maximum and minimum water surface elevations at the breaking position during the passage of one complete wave (Galvin Jr., 1971). As far as a hazard rating is concerned it is the maximum wave height that should be used during the observation period because of its implications for safety. This maximum wave height may change as the tide changes, even when all other weather effects remain constant. It is usually largest at low tide.

Wave Height (m)	0 to 0.49	0.5+	1.0+	1.5+	2.0+	2.5+	3.0+	3.5+	4.0+
General description	Knee High		Waist High		Head High		Overhead		Double Overhead
Wave Height Rating (WHR)	0	1	2	3	4	5	6	8	10

Table 1

3.2 Wave Type Rating (WTR)

Common wave types occurring on Australian beaches include surging, spilling, plunging and plunging with back-blasting. Surging waves are non-breaking symmetrical waves until they reach the water's edge. They have a very small surf zone. Spilling waves break from the top of the crest down the front face of the wave. They dissipate their energy over a large section of the surf zone as they move shorewards. Plunging waves ("Dumpers") break from the top as a plunging jet of water onto the area in front of the wave face. They deliver a lot of energy over a short distance. Plunging waves develop an extreme feature known as back-blasting when the sandbank is very shallow. The front of the wave jets over to such an extent that some water and sand are violently projected out the back of the plunging wave as it hits the sandbank. They are the most hazardous of the four types.

Wave Type	Surging	Spilling	Plunging	Back-Blasting
WTR	0	1	2	3

Table 2

3.3 Wave Period Rating (WPR)

The period is the time between each successive breaking wave in the surf zone. It affects how quickly a competitor can recover stability from one wave to the next, and is especially critical for large craft such as surf boats.

Wave Period	Long (>14s)	9 to 14 secs.	Short (< 9s)
WPR	0	1	2

Table 3

3.4 Zone Length Rating (ZLR)

A surf zone is defined as any region of surf fronts between the outermost breaking waves and the shoreline. Consequently, if there exists a deep channel of water near the shore separated by two rolling turbulent surf fronts, there will be two separate contributions to the ZLR, one for the outer break and one for the shore break.

Zone Length (m)	0 to 1.99	20+	40+	60+	80+	100+	120+	140+	160+
ZLR	0	1	2	3	4	5	6	7	8

Table 4

3.5 Surface Turbulence Rating (STR)

Surface turbulence can be caused by cross-wave interaction from the ground swell and the local wind disturbance acting obliquely to each other. It can also be due to the wind surpassing 17 knots and forming small secondary breaking waves known as chop.

Water Surface	No Cross-waves and No Chop	Either Chop or Cross-waves	Both Cross-waves and Chop
Surface Turbulence Rating (STR)	0	1	2

Table 5

3.6 Littoral Drift Rating (LDR)

When the ground swell direction on open beaches is at an oblique angle to the shoreline, there occurs a sideways littoral drift or drag, which moves the water body parallel to the shore. Sideways movement may also occur from a sandbank towards an outward flowing rip current. Such drags can add to the hazard factor at surf carnivals, particularly when competitions are being held in adjacent sections of the beach.

Latitudinal Movement	No Drag	Low to Moderate Drag	Moderate Drag	Moderate to Strong Drag	Strong Drag
LDR	0	1	2	3	4

Table 6

3.7 Rip Rating (RR)

Outward flowing water movement ("a rip") may be a help or a hindrance to competitors. It is classified here as a hazard because of the risk it affords for the lowest experience-level competitors and craft competitors who lose their craft in a rip.

Longitudinal Movement	No Rips	Mild Rips	Mild to Strong Rips	Strong Rips
RR	0	1	2	3

Table 7

3.8 Other Hazards Rating (OHR)

A hazard rating of 1 is added for additional hazards such as rocks, wrecks, outflow pipes, cold water, floating logs, stingers or the setting sun causing limited visibility.
The rating values are now additively combined to form the Surf Hazard Rating

$$SHR = WHR + WTR + WPR + ZLR + STR + LDR + RR + OHR \quad (2)$$

The values of the SHR will range from 0 to at least 32 based on the maximum and minimum table values, but the range of values may be even higher when multiple zone lengths and other hazards are present. However, there will exist an intermediate ceiling value of the SHR, somewhere inside this range, below which events for all categories may take place with very little risk to any competitors' safety because of the surf conditions. This value will be termed the Acceptable Hazard Rating (AHR) and will be discussed further and used at the end of the next section.

4. EXPERIENCE RATING

The experience of certain groups of competitors to handle the surf conditions at the time of their competition is a governing factor in deciding whether or not to continue at that venue. Within each age and gender group, the skills of the various competitors will vary widely, particularly in local carnivals and heats of State and National Championships. In championship finals, consisting completely of the top competitors, overall skill levels will be much higher. As our primary concern is the safety of all competitors, we therefore suggest at this stage of the development of the model that the level of skills be based on the least-experienced competitors in each category. Even so, all competitors will have passed their annual proficiency test involving surf swimming and board paddling, and this should indicate that they are capable of competing in any surf with a hazard rating less than or equal to a basic acceptable hazard rating (AHR).

4.1 Acceptable Hazard Rating

The surf conditions may often be such that there is potentially little risk in holding a surf carnival at that beach as far as danger to competitors is concerned. To ensure this, an acceptable hazard rating is used as the basic experience level rating for the least experienced competitors in the oldest and youngest categories.
The acceptable hazard rating is based on intermediate fixed values from the seven tables of the SHR. Using wave heights up to 2m, combined surf zone lengths of 60 and 20m these tables yield

$$4 + 1 + 1 + (3+1) + 1 + 2 + 1 + 0 = 14$$

to the AHR.

4.2 Competitor Surf Experience Rating (CSER)

The Competitor Surf Experience Rating should be associated with the most difficult event encountered by that age and gender group, and will vary greatly from the weakest Under 15 and over 60 competitors (rated at 14) to the best Open competitors (rated at 22).

Since they do not compete on skis, the Under 15 competitors' experience will be based on their skills in the board race. This has a higher risk of an accident occurring than in the surf race since the board race has twice the number of objects in the water and half of these (the boards) are a potential hazard when out of any competitor's control. The experience rating for all other age groups will be based on the most difficult craft for that age group. For the purpose of this paper, the order of events from least to most dangerous is swim, board, ski and boat. Two separate tables are presented for Masters' and General Water events respectively using boards or skis, whilst separate tables are presented for Masters' and General Boat race competitors.

The following tables (Tables 8 to 11) list the suggested CSER for various age classes, gender and skill levels. It should be noted here that it would be anticipated that a complete table rating system, based on our initial summarised tables, could be developed to include every water category.

General Water Categories

Category	Limited	Average	Excellent
Under 15 Female	14	16	18
Under 15 Male	15	17	19
Under 17 Female	15	17	19
Under 17 Male	16	18	20
Under 19 Female	16	18	20
Under 19 Male	17	19	21
Open Female	17	19	21
Open Male	18	20	22

Table 8

Masters' Water Categories

Category	Limited	Average	Excellent
30 to 44 Female	16	18	20
30 to 44 Male	17	19	21
45 to 59 Female	15	17	19
45 to 59 Male	16	18	20
60+ Female	14	16	18
60+ Male	15	17	19

Table 9

Boat Events

Boat races have a very high associated risk of accident in the surf because there are five people, five oars and one large craft capable of smashing into each other when the surf boat becomes out of control. Also, the modern surf boat is a sleek racing craft. Stability has been sacrificed somewhat to achieve this design. Consequently, we have given the lowest rating, that of Masters over 200 years, to be equivalent to the lowest under 15 years Female board race and rated all other boat events accordingly.

General Boat Categories

Category	Limited	Average	Excellent
Under 19 and Female Under 23	15	16	17
Male Under 23 and Female Open	16	18	20
Male Open	17	19	21

Table 10

Masters' Boat Categories

Category	Limited	Average	Excellent
Masters over 200	14	15	16
Female masters 120 to 180	14	15	16
Male masters 120 to 180	15	16	17

Table 11

5. COMPETITION SURF SAFETY INDEX (CSSI)

The Competition Surf Safety Index (CSSI) needs to be a relatively easy guide to the risk management decisions that need to be made about whether or not to proceed with certain events at a surf carnival at any particular time or venue or section of a beach.

So, for any given time at a competition venue, the Competition Surf Safety Index (CSSI) relevant to each category is given by

$$CSSI = CSER - SHR \quad (1)$$

where

$$SHR = WHR + WTR + WPR + ZLR + STR + LDR + RR + OHR \quad (2)$$

using Tables 1 to 11.

For any category where the CSSI is negative, it should be recommended that either

(i) these competitors and their cohorts should be moved to an easier beach (with a zero or positive CSSI), or

(ii) their competition be suspended until conditions improve.

References

ABSAMP Australian Beach Safety and Management Program, 1990. http://www.sls.com.au

Galvin Jr., Cyril, 1971, *Waves Breaking in Shallow Water*, in Waves on Beaches and Resulting Sediment Transport (ed. by R.E. Meyer), Proceedings of an Advanced Seminar, University of Wisconsin, Academic Press.

Galton, Barry, 1984, *Gladiators of the Surf*, Reed, Sydney.

Short, A.D. and Hogan, C.L. , 2006, *Rip Currents and Beach Hazards: Their Impact on Public Safety and Implications for Coastal Management,* Journal of Coastal Research, Special Issue No. 12 : Coastal Hazards, pp. 197-209.

Willie Washington of the United States of America ran a 100m time of 13.46 seconds at the age of eight, on 6th of August, 2011.

On the 29th July, 2012, he ran a 12.80 as a nine year old.

This time is 0.2 seconds behind the bronze medal winning time of 12.6 seconds at the 1896 Olympics.

The current record for a twelve year old is Isaiah Green of the United States of America. His time of 11.42 seconds run in 2002 would have seen him win gold in Athens, 1896.

The 10 second barrier was not broken at the Olympics until Jim Hines won in 9.95 in 1968 – and it was not broken again until Carl Lewis won in 9.99 in 1984.

OPTIMAL EXCHANGE BETTING STRATEGY FOR WIN-DRAW-LOSS MARKETS

Darren O'Shaughnessy [a, b]

[a] *Ranking Software, Melbourne*
[b] *Corresponding author: darren@rankingsoftware.com*

Abstract

Since the Betfair betting exchange launched in 2000, sports gamblers have had a gambling forum quite different from the traditional bookmaker. Three features of betting exchanges in particular require new analysis methods extending the Kelly criterion originated by John Kelly (1956):

(i) The ability to lay (i.e., bet against) a team as well as back it
(ii) Negotiation of odds, where one can set one's own odds and wait for another punter to match them, not just accept the market valuation at the time
(iii) The bookmaker takes a fee as a fixed fraction of one's net profit on a market, not as a hidden margin in each betting option's price

In sports where there are more than two possible outcomes, such as soccer (football), usually the prospective gambler will find that if he/she wants to bet on one team using the Kelly criterion, the same criterion will advocate laying against the other team. Basic Kelly betting offers no resolution to these correlated markets, and some punters at traditional bookmakers will instead seek a binary 'handicap' or 'draw-no-bet' market in order to find prices that they can immediately understand.

This paper derives the criterion one should use when investing in a 'win-draw-loss' market, with the important feature that profits are significantly higher by combining back and lay bets than by relying on one or the other. The 'draw-no-bet' approach is shown to be optimal only in a narrow band of cases, where the advantage of having the draw result untaxed outweighs the profits to be gained by effectively backing it.

Keywords: Betting exchange, Kelly criterion, Betfair, Football, Soccer

1. INTRODUCTION

The Kelly criterion (Kelly, 1956) is a vital tool in the armoury of both portfolio investors and gamblers. By maximising logarithmic utility – equivalent to minimising the risk of ruin – Kelly provided the formula that gamblers with perfect probabilistic knowledge must use to grow their bank at the largest expected rate.

With the explosion in sports betting around the world since the rise of the internet, several papers have been written expanding the Kelly criterion, for instance to account for multiple simultaneous independent market investments (Thorpe, 1997), spread betting (Chapman, 2007), and back/lay comparison on betting exchanges (Walshaw, 2010).

Meanwhile, betting exchanges have provided markets that are often more attractive than regular bookmakers, by allowing punters to match their money with peers. Soccer (association football) matches are some of the most popular and therefore most liquid markets, with over $100 million matched on the final of the 2010 FIFA World Cup (Betfair, 2010).

This paper addresses dilemmas that gamblers can feel in sports such as soccer and cricket that have a three-option market. Is it more profitable to back the team that the punter believes is underrated by the market, or match someone else's money by laying the team that appears overrated? And is the "draw no bet" market worthwhile?

2. METHODS

The basic Kelly criterion for a single option on a regular betting market gives the Kelly Bet B as:

$$B = \frac{Mp-1}{M-1} \qquad (1)$$

where M is the team's market price and p is the gambler's presumed probability of the team winning. B is expressed as a percentage of the bettor's bankroll, and a bet should be placed iff $Mp>1$. The formula is derived by maximising *log*(expected bank) with respect to the bet proportion B.

Betfair, which comprises about 90% of the betting exchange economy worldwide (Sydney Morning Herald, 2006), does not build its profit margin into each price like a traditional bookmaker, but instead 'taxes' each market winner on their net profit once the event is resolved. A gambler could have several individual bets on the same market, even arbitraging a guaranteed profit as the odds change, and only pay a fee on his/her net result on the winning option(s). The level of tax t varies from 5% for low-volume gamblers down to 2% for those who have the largest betting history. This leads to an adjusted formula, where M_B is the agreed price for the bet:

$$B = \frac{(M_B-1)(1-t)p-(1-p)}{(M_B-1)(1-t)} \qquad (2)$$

It is immediately obvious that for a single bet, taking Betfair odds M_B is exactly equivalent to taking a slightly lower price at a standard betting shop:

$$M' = 1+(M_B-1)(1-t) \qquad (3)$$

e.g. for a 5% tax, Betfair \$2 is equivalent to traditional \$1.95 while Betfair \$1.20 is equivalent to \$1.19 at a regular bookmaker.

In lay betting, the punter risks $L(M_L-1)$ by accepting a bet of size L from an anonymous peer, having negotiated a price M_L. The Kelly Bet in this case is:

$$L = \frac{(M_L-1)(1-t)(1-p)-p}{(1-t)} \qquad (4)$$

In this paper, we limit our analysis to a combination of betting on one team and laying against the other, ignoring the market price for the central option (the draw). This is done without loss of generality if the market is fully saturated and has automated bet-matching, which is usually the case for Betfair soccer markets. Under this assumption, the draw price can be derived as $M_D = 1/(1-1/M_B-1/M_L)$ but the other two prices capture all necessary market information. We also do not consider the risks and benefits of setting our own odds and waiting for the market to match them, although this should be part of a practical application along with an assessment of the reliability of the punter's presumed probabilities.

To find the optimal combination of bets $\{B,L\}$ we must go back to first principles and maximise W, the log of the expected bankroll. At a traditional bookmaker who offers the equivalent of 'lay' odds (usually called a 'second chance' or 'win or draw' market), this is easily solved.

$$\begin{aligned} W = \; & p_L \log(1-B-L(M_L-1)) \\ & +(1-p_L-p_B)\log(1-B+L) \\ & +p_B \log(1+B(M_B-1)+L) \qquad (5) \end{aligned}$$

Solving for

$$\frac{\partial W}{\partial B}=0 \text{ and } \frac{\partial W}{\partial L}=0 \qquad (6)$$

We find that the maximum bankroll growth is achieved when

$$B_0 = \frac{M_B p_B(M_L-1)-M_L(1-p_L)}{(M_B-1)(M_L-1)-1}$$

$$L_0 = \frac{M_B(1-p_B)-M_L p_L(M_B-1)}{(M_B-1)(M_L-1)-1} \qquad (7)$$

provided both B_0 and L_0 are positive. If only one is positive, the punter should revert to (1).

The situation with tax is more complicated, requiring maximising of the function:

$$\begin{aligned} W_t = \; & p_L \log(1-B-L(M_L-1)) \\ & +(1-p_L-p_B)\log(1+(1-t)^{H(L-B)}(L-B)) \\ & +p_B \log(1+(1-t)(B(M_B-1)+L)) \qquad (8) \end{aligned}$$

where $H(x)$ is the Heaviside step function, indicating that the draw result is only taxed if the lay is larger than the bet.

3. RESULTS

First, consider the simpler situation described in equation (7) where a traditional bookmaker offers odds to bet on a win-or-draw market. In general, these prices tend to be unattractive as the bookmaker has built a substantial profit margin for itself into them; however they are a useful demonstration of the method for the more complex betting exchange situation.

In this example, City is playing United. Our hypothetical punter with a $1,000 bankroll believes that the true probabilities are 60% City wins, 15% United wins, and 25% the match will be drawn. The bookmaker is offering M_B = $2.00 about City, and a "draw or City" market at $1.30, equivalent to laying United at M_L = $4.33.

Using the independent formula (1), the Kelly criterion would advise us to bet 20% or $200 on City. The mean profit is expected to be ($400 × 60% – $200) = $40.

Considering instead the "draw or City" market, the Kelly criterion advises a bet of size $350 (35%), equivalent to laying United for $105 = $350 × (1.3–1). The mean profit is expected to be ($455 × 85%– $350) = $36.75.

To reconcile these two criteria, we must use formula (7) to find $\{B_0 = 0.13571, L_0 = 0.06429\}$. I.e., bet $135.71 on City and simultaneously bet $214.29 on "draw or City" (equivalent to laying United for $64.29). For an exposure of $350 – in this case the same as "draw or City" alone – the punter has increased his expected profit to $49.64, or 14.2% of his outlay.

Paradoxically, it should be noted that the punter has effectively taken a price on the draw outcome that the Kelly criterion would advise has an expected loss. The punter believes that the draw is a 25% prospect, but the difference between the odds of $2 (50%) and $1.30 (77%) is 27%, meaning that he is paying a premium for including this option in his betting portfolio. Additionally, if the match results in a draw he will still suffer a net loss of $71.43. However, as the goal is to minimise the risk of long-term ruin, the increased diversification to include the draw is the correct strategy.

The Betfair version of this problem in equation (8) must take into account three different possibilities:

i. $L > B$, so the draw will be a net loser
ii. $L < B$, and the net profit from a draw will be taxed
iii. $L = B$, "draw no bet"

The log formula to be maximised is different in all three cases, so the zeroes in the derivatives must be examined for domain relevance and compared with each other.

3.i. The Case L > B

Solving (6) for W_t in (8) with the taxed draw gives:

$$B_1 = B_0 \frac{M_L - t}{(1-t)M_L} \tag{9}$$

$$L_1 = L_0 + \frac{t(M_B p_B - M_L^2 p_L (M_B - 1) + M_L(1-p_L) + M_B M_L (p_L - p_B))}{(1-t)M_L((M_B-1)(M_L-1)-1)}$$

provided $L_1 > B_1$.

3.ii. The Case L < B

Solving (6) for W_t in (8) with the untaxed draw gives:

$$L_2 = L_0 \frac{(1-t)M_B + t}{(1-t)M_B} \tag{10}$$

$$B_2 = B_0 + \frac{t(M_B(1-p_B) - M_B M_L (1-p_B) + M_L p_L)}{(1-t)M_B((M_B-1)(M_L-1)-1)}$$

provided $L_2 < B_2$.

3.iii. The Case L = B ("draw no bet")

By eliminating the draw outcome, (6) is simply solved for the first and third terms of (8) with B set to L:

$$B_3 = L_3 = \frac{(1-t)M_B p_B - M_L p_L}{(1-t)M_B M_L (p_B + p_L)} \tag{11}$$

This is the common boundary of the other two cases.

Examples

Returning to our City vs United example, consider a market where the exchange prices are M_B = $2.05 about City, and M_L = $4.50 for United. For an individual market option, this is equivalent to the standard bookmaker's prices earlier in this section after a t=0.05 tax is factored in.

Checking the three case functions, case ii is consistent and outperforms case iii, which dominates case i along the entire boundary (the maximum of W_t occurs outside of the $L > B$ domain). The formula recommends values $\{B_0 = 0.15837, L_0 = 0.04266\}$. I.e., bet $158.37 on City and simultaneously lay United for $42.66 (lay exposure $149.30, total exposure $307.67). The expected net profit on the market is $44.02, or 14.3% of his outlay. His loss in the case of a draw is $115.71. Using the exchange tends to push the successful strategy in the direction of betting on the favourite as opposed to laying against the underdog.

Optimal Back/Lay From $1,000 bankroll					
City	Draw	United	Case	Back (B)	Lay (L)
$50	$11.70	$1.12	L>B	$9	$498
$20	$6.75	$1.25	L>B	$21	$350
$10	$4.65	$1.45	L>B	$41	$239
$7.50	$4.10	$1.60	L>B	$53	$197
$6.00	$3.75	$1.75	L>B	$66	$166
$5.00	$3.50	$1.95	L>B	$79	$140
$4.00	$3.35	$2.20	L>B	$98	$111
$3.50	$3.25	$2.46	L=B	$104	$104
$3.35	$3.25	$2.55	L=B	$104	$104
$3.20	$3.25	$2.62	L=B	$103	$103
$2.90	$3.20	$2.90	L<B	$114	$90
$2.60	$3.25	$3.25	L<B	$128	$76
$2.40	$3.25	$3.60	L<B	$139	$67
$2.20	$3.35	$4.05	L<B	$152	$57
$2.00	$3.50	$4.70	L<B	$169	$47
$1.80	$3.70	$5.75	L<B	$190	$36
$1.60	$4.10	$7.60	L<B	$219	$26
$1.40	$5.00	$11.50	L<B	$259	$15
$1.20	$7.80	$26	L<B	$329	$6
$1.10	$13.50	$62	L<B	$391	$2.20

Table 1: Effective Strategy for a Range of Markets

To examine how the optimal back/lay proportion changes with varying odds, a series of price sets was generated from the cumulative normal distribution with a fixed z difference of 0.3 between the market odds and the punter's probabilities. The central 'draw' option was given a width of 0.8 on the z scale to mimic real soccer draw odds in professional leagues. For example, the market centred on z=0 would have City and United both on a price of $2.90 (equivalent to 34.5% probability, or $z < -0.4$). The punter's belief is that City has a 46.0% chance of winning ($z < -0.1$), compared to 24.2% for United ($z > 0.7$). While this method produces superficially credible sets of odds, a more precise simulation of soccer should use an accepted modelling approach such as that recommended by Dixon (1998).

Table 1 shows the optimal strategy for a range of odds, and Figure 1 plots the function, clearly showing a 'kink' for the narrow range of situations where the gambler should attempt not to make a profit on the draw, in order to avoid tax on that result. The function using standard bookmaker odds does not display such a discontinuity in the derivative.

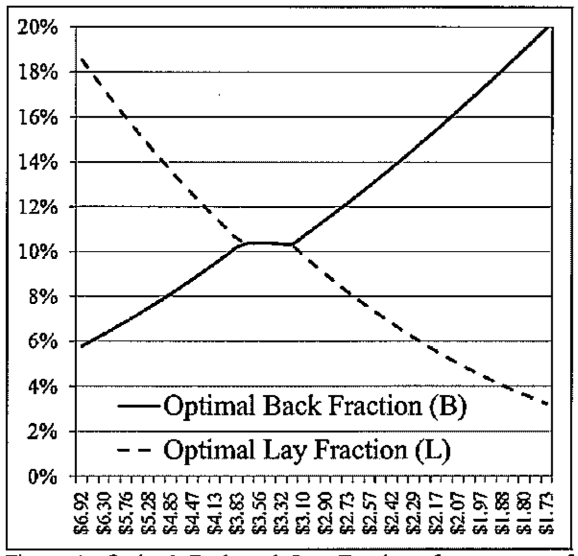


Figure 1: Optimal Back and Lay Fractions for a range of favourite's prices, showing a narrow central zone where 'draw no bet' is the optimal strategy

4. DISCUSSION

Betting via an exchange has subtle and surprising repercussions for optimal gambling strategy. The way in which a bookmaker profits is quite different from the situation at an exchange like Betfair, which does not set the odds centrally but takes a fraction of each payout when the market settles.

This paper has extended the well-known Kelly criterion to the popular sport of soccer, and shown that use of this advanced strategy leads to a substantial increase in expected profit – greater than 20% in our City vs United example.

Naturally, punters in the real world ought to revisit the assumptions of section 2, particularly if the market is not as liquid as they might wish. In practice, blindly using the full Kelly Bet fraction would be a risky rollercoaster for most punters, as they do not account for the error in the probabilities generated when they frame the market. A more complete approach might use a Bayesian distribution of predicted market outcomes, taking into account a variety of known and unknown factors in the game then optimising a more complex log-utility function.

A handy rule of thumb from the equations is that in general, a punter who wants to back the favourite should put most of his/her money into that option, while backing the underdog is usually not as efficient in growing the bankroll as laying against the favourite – particularly when the favourite's odds are close to 50/50.

5. CONCLUSIONS

This paper has solved the Kelly criterion for the case of win-draw-loss markets. This is immediately applicable to soccer, cricket, and other sports with a reasonable likelihood of neither team winning such as hockey and chess.

It is also immediately applicable to sports such as Australian Rules Football, where the punter must decide how to allocate his funds to a head-to-head (win/loss) or 'line' bet, which are the two most liquid markets. The range of outcomes between zero and the published line handicap can be treated as the middle outcome in the formulas published here.

More generally, future work could extend the methodology to n published lines and numerically find the optimal W_t for a betting portfolio that would potentially be spread across a number of them.

References

Chapman, S.J. (2007). The Kelly criterion for spread bets. *IMA Journal of Applied Mathematics, 72*, 43-51

Dixon, M.J. (1998). A Birth Process Model for Association Football Matches. *The Statistician, 47*, 523-538

Kelly, J.L. (1956). A New Interpretation of Information Rate. *IRE Transactions on Information Theory, 3*, 185-189

Sydney Morning Herald (2006). Betfair out to conquer the world, February 6, 2006. *Available at http://www.smh.com.au/news/business/betfair-out-to-conquer-the-world/2006/02/05/1139074110158.html*

Thorp, E.O. (1997). The Kelly Criterion in Blackjack, Sports Betting, and the Stock Market, in *Finding the Edge: Mathematical and Quantitative Analysis of Gambling (eds. Eadington, W.R. & Cornelius, J.A.)* 163-214

Walshaw, Rowan. (2010). Kelly Criterion – Part 3a – Backing and Laying Bets with Betfair. *Available at http://www.aussportsbetting.com/2010/07/18/kelly-criterion-backing-andlaying-bets-with-betfair/*

DATA MINING IN SOCCER: PERFORMANCE PATTERN CLUSTERING OF TOP NATIONAL TEAMS

Abbas Saghaei[a], Elaheh Gholamzadeh Nabati[a], Parvin Alavi[b]

[a] *Department of Industrial Engineering, Science and Research branch, Islamic Azad University, Tehran, Iran.*
[b] *Central Tehran Branch, Islamic Azad University, , Tehran, Iran.*
Emails: a.saghaei@srbiau.ac.ir, e.nabati@srbiau.ac.ir

ABSTRACT

Team performance study is an important issue in sport research. It is essential for sports organizations to identify if their teams are performing satisfactorily. In this research, a new method for analyzing the performance data of soccer teams is presented. The method is based on data mining and statistical quality monitoring. We use data mining techniques to develop a team performance monitoring framework. For this purpose, we apply functional data clustering to discover patterns and characterize different types of trends. The input data are in the form of temporal curves. The temporal performance curves show the monthly ranking scores of teams which have been published by FIFA during a period of 2006- 2010. The curves for 67 national soccer teams are obtained. We consider the performance curves as a measure of quality for each team, which can be monitored over time. By statistical monitoring, the teams that go through unusual trends are identified. Our method provides interesting output that might be applicable in sports practically and additional information for soccer managers and coaches, which may help them in identifying changes and enables them to improve their team's performance.

Key words: Soccer, Functional Data Clustering, Team Performance, Data Mining, Statistical Quality Monitoring

1. INTRODUCTION

Team performance study is an important issue in sport researches. It is essential for sports organizations to identify if their teams are performing satisfactorily. Sport organizations are investigating new ways to improve their methods of training and management. With the increasing amount of sport data stored in databases, there is a need for organizations to develop new methods in order to extract knowledge from these data. The obtained information can be useful in improving training activities and for the managers in making their decisions.

We observe an increase in the number of articles utilizing data mining methods in sports. Data mining is an interdisciplinary area of statistics, artificial intelligence and database science. Its goal is to extract the useful information from huge amounts of data. Data mining methods can be classified into supervised learning and unsupervised learning methods.

Supervised learning methods or learning by example, includes methods in which a set of input variables of a data set is used to predict the response variable(s). Neural Networks (NNs), regression, linear discriminant analysis, and classification are among the supervised methods being used in sport researches. David, Pasteur, Saif Ahmad & Janning (2011) applied Neural Networks for predicting the result of NFL games. McCullagh (2010) proposed NNs as a useful tool to selection of players in the annual Australian

Football League. He stated that this method has the potential to assist managers in identifying talented players. Fischer, Do, Stein, Asfour, Dillmann & Schwameder (2011) compared the ability of Neural Networks with Support vector machines on modeling the kinematic patterns during walking and running. European soccer team A.C Milan can predict the probability of injury for each player by using data mining methods (Solieman, 2006). Heiny & Blevins (2011) applied discriminant analysis to predict the offensive play calling of the Atlanta Falcons Football team during the 2005 season. Mills & Salaga (2011) used Random Forest for analyzing voting patterns in rating baseball players. Levernier & Barilla (2007) applied a logit model in the field of baseball. They proposed four logistic regression models to predict the probability of winning for a team. The impact of several factors was analyzed in this model.

Unsupervised learning deals with inferring the properties of a data set and finding hidden patterns in the data when there is no response variable (Hastie, Tibshiran, & Friedman, 2009). Among the unsupervised learning methods, clustering and self-organizing maps (SOM) have been used in sports data mining. Self organizing maps were used to group movement patterns of basketball players when shooting from different distances (Lamb, Bartlett, & Robins, 2010). Pfeiffer & Perl (2006) used DyCoN, a special type of SOM, for analyzing the tactical structure of handball games in Women's world Championship 2001. Perl & Memment (2011) applied SOM in net-based game analysis. They attempted to find offensive and defensive patterns based on the position of players on the field.

Clustering is one of the most useful unsupervised learning approaches, but there isn't much literature on the application of this technique on sport data. In sport video analysis, clustering is applied to group scenes and patterns. In swimming, Chen, Homma, Jin & Yan (2007) used hierarchical agglomerative clustering to identify elite swimmers race patterns in four swimming championships. They analyzed four performance criteria (swimming speed, stroke length, turning speed and time). Chen, Chen, Jin & Yan (2008) presented another study on swimming race patterns on a larger data set of both male and female athletes and clustered them by swimming speed and stroke length. Puterman & Wang (2011) modeled competitiveness in a basketball league by dynamic clustering. They developed a model in which teams with good performance at the end of the season, are promoted to higher clusters (divisions) and teams that perform poorly, relegate to lower divisions. In the last three studies, a classic kind of clustering is applied. In which the input data is in the form of a vector. In this paper we apply a new kind of clustering, Functional data clustering (Ramsay and Silverman, 2005), in which the data units are functions or curves defined on some interval, rather than just one observed data point in the interval. Analysis of curve data has the advantage of using measurements over time (including length of intervals) instead of utilizing summary statistics.

This method is recently developed and gained significant attention in a variety of disciplines, for example, Genetics (Duan, Liou, Shi & Didonato (2003), Song, Lee, Morris & Kang (2007)), Neurology (Jin, Ham & Kim, 2005) and engineering (Goia, May & Fusai, 2010). The method proposed here, has a further advantage. It consist of a trimming procedure that allows us to seperate curves with uncommon patterns from the rest of patterns. Trimming procedures cuts the outliers in data and makes the clustering more robust. Thus, the out-put clusters are more accurate than classical clustering. In this paper we want to take advantage of this possibility, recognize behavioral patterns of team performance and identify the teams with homogenious patterns. We use the FIFA World Rankings for men's national Soccer teams, during the time period between July 2006 to May 2010. The scores given by FIFA to each team have been used to obtain the performance curve. We consider the performance curves as a measure of quality of team, which can be monitored over time. Coaches usually compare performance of their team over few points in a specific time interval and don't consider the issue over a long time period. Considering the curves of performance can provide a good insight into the status of team play and may help in finding probable causes of its change. Our method is a knowledge discovery approach and explores new information from the teams' performance data. This method can provide coaches and sports managers with extra information for performance analysis, enables them to make comparisons, discover the causes of change and implement suitable actions to improve their team's performance. This paper is organized as followed: Section 2 discusses the FIFA performance ranking procedure. Section 3 provides the methodology of our research. Section 4 is dedicated to the findings and results. Section 5 provides a discussion of the results. And finally, Section 6 presents concluding remarks for this study.

2. FIFA PERFORMANCE SCORING PROCEDURE

In order to have a comparable and reliable quality measure, we use the information published by FIFA, whom provide a monthly rankings of national soccer teams. According to the FIFA Factsheet (FIFA, 2011) the scores in FIFA performance rankings are calculated according to the equation below:

$$P = 100(M.I.T.C) \quad (1)$$

Where M is the number of points that teams can obtain in each match (three points for a victory, one point for a draw and zero points for a loss). I shows the importance of the match and has the value between one and four. The value increases as the importance of match rises; one point for the friendly matches and four points for world cup final competitions. T indicates the strength of the opposing team and is calculated through the following formula:

[200 – Ranking position of opposition] ÷ 100 (2)

And C is the strength of confederation and the values are between 0.85 to1.

3. METHOD

This section introduces briefly the framework of our research. The steps taken were: data collection and transformation, trimmed K-means clustering and analysis of results. The performance scores of the top 100 teams were collected. However, as 23 teams were not constantly among the top 100 teams during the selected time interval (July 2006, may 2010), they were omitted. Eventually, 67 teams remained. Data preparation tasks were done before cluster analysis. The data were normalized and spline basis functions were applied on curves to interpolate and reduce the dimension of each curve (Ramsay and Silverman, 2005). Splines basis is a set of independent piecewise polynomials which is used to interpolate curves and has the characteristics that we can model functions with. For additional information about spline bases and curve clustering see Garcıa-Escudero and Gordaliza (2005). We applied B-spline, a popular type of spline, with 5 knots. This number of knots was sufficient for our purpose of analysis. It should be noted that, in this paper we applied splines for the purpose of smoothing the curves. Other applications, such as prediction, are not in the scope of this paper.

3.1. TRIMMED K-MEANS

K-means is a cluster analysis technique that partitions data to K groups. Cluster membership is determined by calculating the center for each group and assigning each data point to the group with the closest cluster center. K-means is sensitive to outlier data. To makes the algorithm more robust, Garcia-Escudero and Gordaliza (2005) proposed a procedure to trim α percents of observations. The procedure is called "impartial trimming"; which suggests that, we specify the value of α, but the curves which are going to be trimmed will be determined by the data itself. As such, the trimmed observations are usually the data points which are unique from the rest, and can be consider as outliers. Afterwards, the K-means algorithm steps are implemented to the remaining points in the data set. K-means clustering with impartial trimming can be formulated as the following:

$$\min_{S} \min_{m1,\dots,mk} \frac{1}{[n(1-\alpha)]} \sum_{K=1}^{k} \sum_{x_j \in S} \|X_j - m_k\|^2 \quad (3)$$

Where S is a subset of curves, contains [n(1-α)] data curves as $X_1, X_2, \dots, X_{n(1-\alpha)}$. (Here, the [.] shows the integer part of the number). This problem allocates each curve to one of k clusters in a way that the distance to cluster centers, m_j ($j=1, 2, \dots, k$) is minimized. The parameter α, is a number belonging to interval [0,1] and n(1-α) is the number of curves remaining after trimming (Garcıa-Escudero & Gordaliza, 2005). In this paper we assume that 0.05

percent of curves should be trimmed off. We examined different trimming sizes on our data and realized that α=0.05 is appropriate. This trimming procedure increases not only the accuracy of clustering but also provides a basis for further investigation on the trimmed observations and the possible causes of their difference.

3.2. CONCEPT OF QUALITY CONTROL IN SPORT

We chose trimmed K-means for three reasons: first, it is applicable to functional data (curves). Second, trimmed k means is more robust than classical k-means. Third, the trimming procedure is similar to removing outlier data in a statistical quality control. Statistical quality control is concerned with monitoring performance of a process over a time using measurements of quality characteristics. Elimination of outliers reduces the process variation and provides more stable estimates of process performance. In this paper, we use the concept of "process monitoring" to propose a framework for monitoring the status of teams over time. This concept is new in both statistical quality control and sport research.

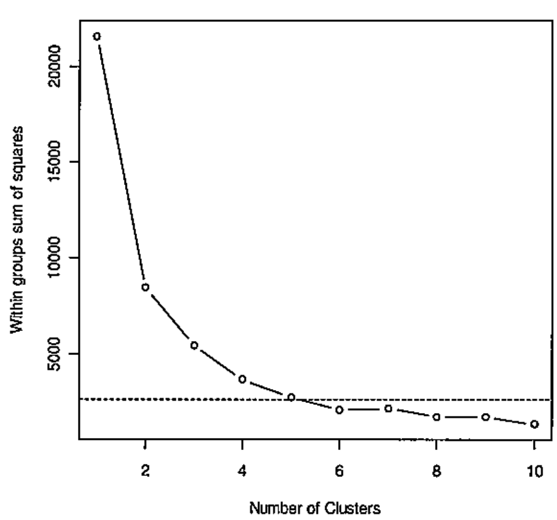


Figure 1: Within cluster error sum of Squares

3.3. NUMBER OF CLUSTERS

To choose the number of clusters, we calculated the with-in cluster sum of squares error (Fig. 1). We chose the model whose total with-in cluster variation is less than the 0.5 standard error. Figure 1, shows the with-in cluster dispersion for different cluster numbers. The Dotted line indicates the 0.5 standard error. The first point in which the minimum with-in cluster dispersion occurs is six clusters (K=6). More clusters, have no significant impact on the reduction of variance and doesn't provide a better model of the data. We assume that six patterns exist within our curves.

4. RESULTS

In this section the results we obtain by applying trimmed K-means clustering (see section 3.1) on the performance curves are presented. We have conducted two different studies; first, clustering with a trimming procedure and second, clustering of performance curve trends. The first one is done on the teams by considering their rank in the FIFA table and the other has been done on the trend of changes in the scores after eliminating the effect of rank on the team's curve.

4.1. CLUSTERING BASED ON TRIMMING

The output was 6 groups of performances and four trimmed observations. Figure 2 and Table 1 shows the trimmed curves and obtained clusters. The trimmed observations diagnosed are Spain, Brazil, Czech Republic and Nigeria teams. One of the advantages of using trimmed K-means algorithm is the capability of this algorithm to exclude the data that has less similarity to the rest. The study of these trimmed observations leads to interesting information. By inspecting several trimming sizes, we found an alpha of 0.05 to be the most suitable for our data. We increased the amount of alpha (trimming size) from zero to 0.5. At a small trimming size, alpha=0.02, just one team was trimmed (Spain). At alpha=0.05 four teams were trimmed and they were Spain, Brazil, Nigeria and Czech Republic. We studied the behavior of these teams. Spain, Brazil and Nigeria are the first three trimmed teams; having the most different performance patterns compared to other 63 teams.

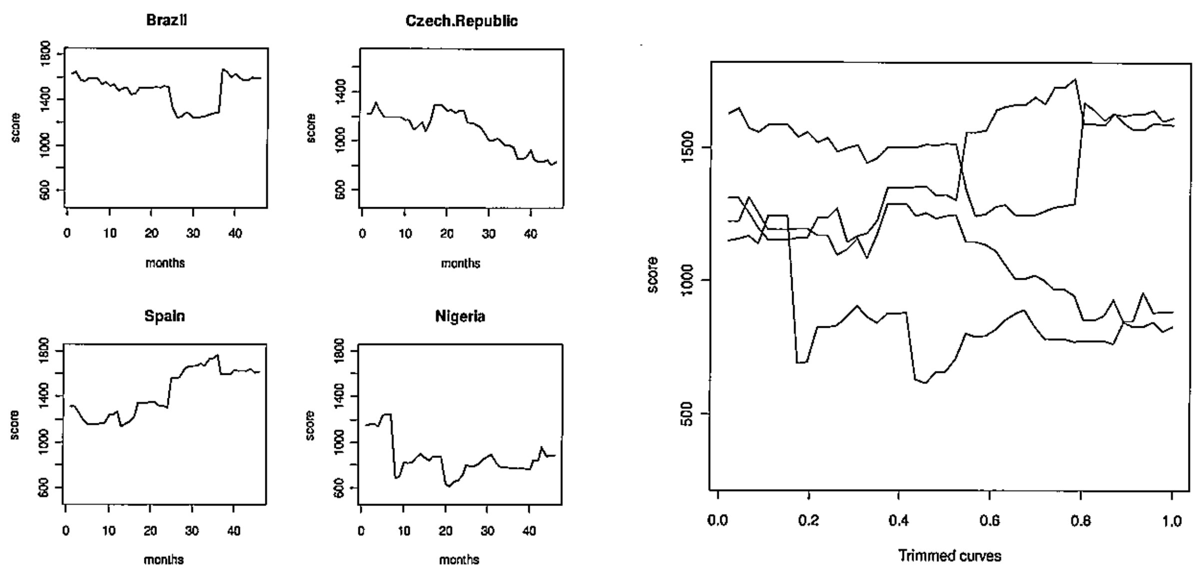


Figure 2: Trimmed curves. X-axis shows the time and Y-axis shows the score points.

Spain is the only team among the top ten teams that has a sharp increasing trend. The other top ten teams have a decreasing trend. Brazil has different performance pattern: decreasing at first, followed by two significant shifts (see Figure 2). Brazil was world champion for 5 years (2001-2006) but they lost first place following a defeat against Portugal in January 2007. Brazil had a decreasing trend afterwards. In July 2009, Brazil won the FIFA Confederations cup; promoting its ranking score by 384 points, creating an upward shift in its performance curve (Figure 2). Nigeria is the third trimmed curve. This team had a deep fall (downward shift) which dropped its position 27 places in FIFA world rankings. Nigeria's significant drop has made it different from all other teams (Figure 2). The Czech Republic shows a decreasing trend, gaining an average rank of 12 during the time period of this study. Other teams with average ranks between 10 and 20 are in cluster 2 (table 1).

The non-trimmed curves were clustered into six groups. The divisions are provided in Table 1. In order to analyze the characteristics of each team, we had to determine which factors play an important role in better functioning (or malfunctioning) of the teams.

Some of these factors are:

A. Having a good coach.
B. Having top players (usually they are those who have been chosen as the best player of the year in World cup or European championship, Asian Nations Cup, African Nations Cup, Copa America and the Gold cup).
C. The achievements of the team, i.e. qualifying for World Cup matches and obtaining good results or intercontinental matches (each federation).

We gathered and studied the data concerning:

I. The number of times the coach of the team had been changed in this specific portion of the time.
II. Qualified players of each team.
III. Achievements of the team in 2006 to 2010.

- In table 1, cluster 1 consists of the teams like Argentina, England, France, Italy and Germany. All these teams are among the top ten rankings of the world. It could be seen that the number of top players in these teams are more than the other clusters and all these teams have had at least one top-player in the studied period of time.

  On the other hand, there's been no or minimal change of coaches in the mentioned time. For example: the French team had the same coach for the entirety of this period. Similarly, the Germans have changed their coach just once.

- The second group, are the teams with an average rank position from 20 to 35. These teams have had less top players than the first group (with a maximum of one or two players) and we observed that there have been more coach changes in their teams.
- In the third group the teams show more variation in the rankings (they ranked twentieth to fifty-sixth), coach changes are reported up to 4 times. The maximum number of top players in this group reaches five (for African Ghana and Cameroon). This number is interesting, but we have to consider that, these two teams have had a good number of top players in the CAF awards 2006 to 2010.
- The group four and group five teams have less top players, and experience more coach changes.
- In group six, you see the two teams of Qatar and Iraq placed in the same cluster, by having two top players each (on the Asian nation's cup). As a result, we observe that more powerful (or more qualified teams) experience minimal change of coaches and management. Their ranking in the FIFA table has less variation and more stability. Also, they benefit from more top players than the other teams.

	Country	Continent	Avg Of position	total Number best players	coach change	championship continental-confederations
Trimmed	Brazil	South America	2.5	7	1	5th FIFA2006+qualified FIFA2010+1st confederations'*2 and 1st Copa America
	Spain	Europe	4.4	6	1	FIFA2006 qualified+1st FIFA2010+1st Euro2008
	Czech Republic	Europe	12.2	2	3	qualified FIFA2006+ qualified Euro 2008
	Nigeria	Africa	24.8	0	5	FIFA 2010 qualified+ 3rd CAF cup
cluster 1	Argentina	South America	4.5	10	2	2nd Copa America+ FIFA2006+qualified FIFA2010
	England	Europe	8.8	4	2	7th FIFA2006+qualified FIFA2010
	France	Europe	7.2	1	0	2d FIFA2006+qualified euro2008+qualified FIFA2010
	Germany	Europe	4.6	1	1	3rd FIFA2006+3rd FIFA2010+2d Euro2008
	Italy	Europe	3.1	3	1	1st FIFA2006+qualified FIFA2010+8th euro 2008+5th confederations cup
	Netherlands	Europe	5.5	4	1	FIFA2006 qualified+2nd FIFA2010
	Portugal	Europe	8.9	6	1	4th FIFA2006+qualified FIFA2010+qualified Euro2008
cluster 2	Australia	Australia	34.9	1	3	qualified FIFA2006+qualified FIFA2010+7th AFC cup2007
	Chile	South America	35.1	1	1	0
	Egypt	Africa	29.3	1	0	1st CAF cup*3
	Greece	Europe	14.9	0	0	qualified FIFA 2010+qualified Euro2008
	Israel	Asia	28.1	0	1	0
	Mexico	North America	19.8	2	3	FIFA2006 qualified +FIFA 2010 qualifed+1st Gold cup 2009+2d Gold cup 2007
	Paraguay	South America	25.0	1	0	FIFA2006 qualified+8th FIFA 2010+ qualified Copa America
	Russia	Europe	17.5	0	0	3rd Euro 2008
	Serbia	Europe	27.0	0	3	FIFA 2006 qualified+ FIFA2010 qualified
	Switzerland	Europe	27.0	0	1	FIFA2006 qualified +FIFA 2010 qualified+euro2008 qualified
	Turkey	Europe	22.6	0	1	3rd or 4th euro 2008
	Ukraine	Europe	20.5	0	2	8th FIFA2006+qualifed euro 2008
	Uruguay	South America	22.4	0	1	4th FIFA2010+4th Copa Ameica 2007
	USA	North America	20.8	1	0	FIFA2006 qualified+FIFA2010 qualified+1st Gold cup 2007+2d Gold cup 2009
cluster 3	Belgium	Europe	56.5	0	3	0
	Bulgaria	Europe	27.5	0	4	0
	Cameroon	Africa	16.9	5	4	2d CAF cup +CAF quarterfinal*2
	Colombia	South America	32.3	0	3	0
	Croatia	Europe	10.8	0	1	5th Euro 2008+ qualified FIFA2006
	Denmark	Europe	27.7	0	0	qualified FIFA2010
	Ghana	Africa	28.8	5	2	qualified FIFA 2006+7th fifa2010+3rd place CAF cup
	Norway	Europe	39.3	0	0	0
	Poland	Europe	32.3	0	2	FIFA2006 qualified+euro2008 qualified
	Romania	Europe	20.3	0	1	Euro 2008 qualified
	Scotland	Europe	25.9	0	3	0
	Sweden	Europe	27.2	0	1	FIFA2006 qualifed+euro2008 qualified
cluster 4	Costa Rica	South America	55.0	0	3	3rd or 4th Gold cup 2009+ FIFA2006 qualified
	Ecuador	South America	41.7	0	1	qualified FIFA2006+pan games champion
	Honduras	North America	48.1	1	1	3rd or 4th Gold Cup2009+qualified FIFA2010 for the first time
	Republic of Ireland	Europe	38.3	0	2	0
	Slovakia	Europe	45.5	0	2	FIFA 2010 qualified for the first time
	Slovenia	Europe	66.0	0	1	FIFA2010 qualified
cluster 5	Finland	Europe	48.1	0	1	0- but good coach
	Hungary	Europe	55.2	0	2	0
	Iran	Asia	47.4	1	2	FIFA2006 qualified + qualified AFC cup 2007
	Japan	Asia	39.2	2	2	FIFA2006 qualified+ FIFA 2010 qualifed+4th AFC cup 2007
	Korea Republic	Asia	48.6	0	1	FIFA2006 qualified+ FIFA 2010 qualifed+1st east Asian cup2008 +3rd AFC2007
	Mali	Africa	47.7	1		0
	Morocco	Africa	46.6	0	5	African cup qualified *2
	Northern	Europe	39.5	0	1	0

	Ireland					
	Tunisia	Africa	46.0	0	3	FIFA 2006 qualified+CAF2007 qualified*2
cluster 6	Canada	North America	70.9	0	4	3d gold cup
	Macedonia	Europe	58.7	0	1	0
	Iraq	Asia	79.8	2	7	1st AFC cup2007
	Lithuania	Europe	62.7	0	2	
	Oman	Asia	82.7	0	4	AFC cup 2007 qualified
	Panama	North America	70.5	0	3	qualified Gold cup
	Peru	South America	67.8	0	3	Copa America qualified*3
	Qatar	Asia	82.6	2	2	qualified AFC 2007
	Saudi Arabia	Asia	58.7	0	3	FIFA 2006 qualifeid+2rd AFC 2007
	Uzbekistan	Asia	65.8	1	2	qualified AFC cup
	Wales	Europe	66.0	0	0	0
	Zambia	Africa	71.2	0	1	qualified African cup 2010+2008

Table 1: Member countries of each cluster and trimmed observations

4.3. TREND-BASE CURVE CLUSTERING STUDY

In this section we present the results obtained by applying trimmed K-means clustering (see section 3.1) on the performance curves. We conduct clustering on pure trends without the effect of the y-intercept. For this purpose, we subtract the first score point from the rest of the points for each team; this resembles to shifting the curve in the direction of the Y-axis, in a way that no change in the shape of curve occurs. Our first aim was to answer the question: does any homogenous behavior exist between teams regardless of where the teams are located in the chart? The second aim was to determine how is the performance of each team changing, as well as to answer the question: does a team show growth in its performance? Although clustering is not necessary to learn about individual performances, it would be time consuming in our case that the number of teams is high. Figure 3 and Table 2 show the results. In Cluster one, the Netherlands (who are among the top charted teams) shows a noisy moderate decrease that starts to increase at the end points. Some other teams, like Ukraine, Iran & Morocco, from the middle of the chart also exhibit the same behavior. Teams from cluster two show an initial decline, proceeding with a constant behavior. England, Denmark, Colombia and Tunisia follow this behavior. This trend shows that the performance of these teams have not been improved during this time period when compared to their pervious performances. Performance among cluster three members were initially rising but remained constant afterwards. Cluster four shows a growing trend suggesting that these teams have improved their own performance with a growth being compared to their previous performance records. A good example of this group is the USA team which qualified for both 2006 and 2010 world cups and did not change its coaching board. Members of cluster five are just three teams, all from the top chart, Argentina, France and Italy which had a sharp decrease in their performance scores. Cluster six members are going through a mild but continues growth. No team from the top 10 has been placed in this cluster. Variation within cluster was calculated and shown in table 3. Cluster two has the least variation among the six. That is, the teams in cluster two are more likely to be homogenous and follow the same behavior.

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Belgium	Cameroon	Croatia	Canada	Argentina	Australia
Macedonia	Colombia	Finland	Costa Rica	France	Bulgaria
Germany	Denmark	Greece	Honduras	Italy	Chile
Ghana	Ecuador	Hungary	Panama		Egypt
Iran	England	Israel	Paraguay		Iraq
Morocco	Peru	Mali	Republic of Ireland		Japan
Netherlands	Portugal	Northern Ireland	USA		Korea Republic
Oman	Switzerland	Russia	Uruguay		Lithuania
Poland	Tunisia	Saudi Arabia	Wales		Mexico
Qatar	Uzbekistan	Scotland			Norway
Romania					Serbia
Sweden					Slovakia
Turkey					Slovenia
Ukraine					Venezuela
Zambia					

Table 2: Member countries of trend-based study.

To verify the accuracy of our proposed procedure in section 4 and to check how much information our results provide, we can compare the performance of the teams in the FIFA World Cup 2010 competitions with the expected performance that our results suggest. Our study covers the time period up to June 2010 which was the start of the South Africa World Cup competitions. Our results suggest that Spain, one of the trimmed observations, was the only team among top charted teams that had an increasing trend. Therefore, from our findings, it had the chance of gaining good results in the FIFA World Cup 2010; which was the actual outcome of the event. Netherlands gained the third place in this competition, as shown in figure 3 and table 2. This country is located in cluster one which had a mild decreasing trend, beginning to improve at the end points. Gaining third place supports our findings that the teams in cluster one (Figure 3) were improving. In our results, Argentina, France and Italy are the members of cluster 5 (Fig. 3) with a sharp decrease. This trend shows a little likelihood for these teams to exhibit high performance (although they are strong competitors and are among the top ten teams in the FIFA ranking) and the results of world cup competitions also suggest this. Finally, consider Uruguay, who showed great performance during FIFA 2010. This follows our expected results, with a cluster 4 member, increasing in its performance.

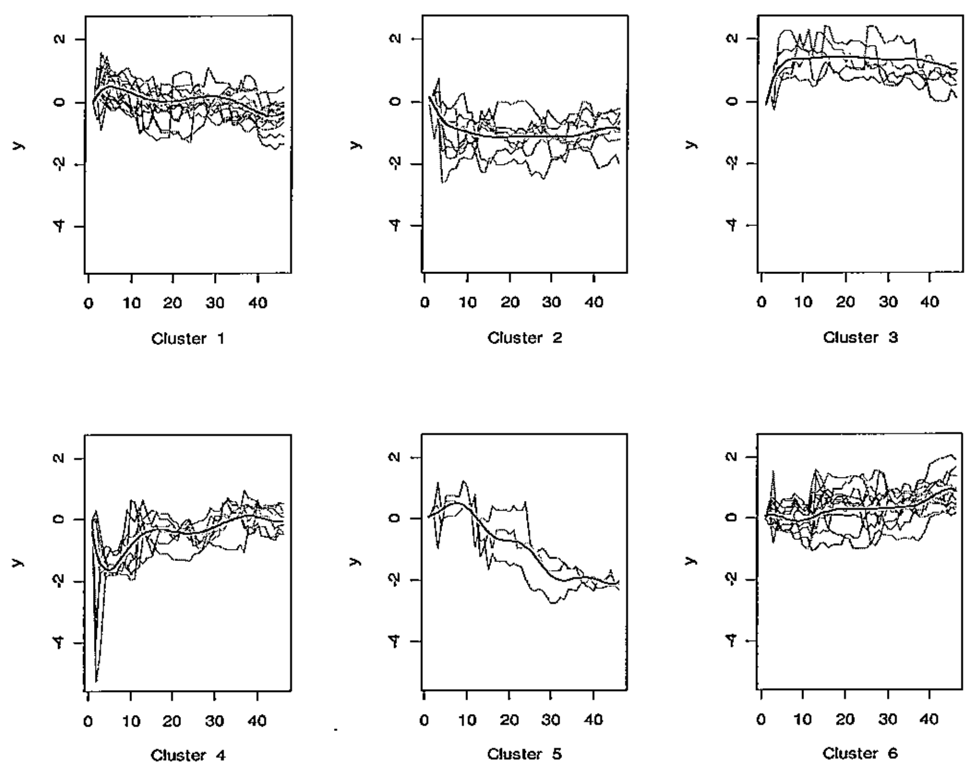


Figure 3: Trend-based clustering Study. X-axis shows the time interval (months) and Y-axis shows the scaled score points.

Cluster variance					
Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
245.9996	134.6354	310.4851	421.8452	461.1653	617.3867

Table 3: Within cluster variation is used to decide how homogenous the clusters members are.

5. DISCUSSION

Our approach can provide a novel method and classified information when the number of the teams is high. Like any other method, there are some advantages and disadvantages that should be considered.

1. Curves in each cluster follow the same behavior, suggesting the trimmed K-means method doesn't allow a team with a decreasing behavior to be allocated with an increasing team in the same cluster. In order that the average show a constant behavior. This issue can be observed in Figure 3 (trend-based analysis).
2. There is no specified way of determining trimming size as well the cluster numbers. However, by testing different trimming sizes and cluster numbers, this problem can be alleviated. Selection of both parameters depends on the level of precision for which the sport managers and analysts are interested. For example, the number of clusters can be defined fewer or more. The more the number of categories, the more specific the level of details we can achieve.
3. One advantage of our method is that, it provides a big picture of the status of teams. In sport activities, a team is usually compared with one or a few other rival teams. But the proposed method has the capability to show more than 50 teams together in a long time period.
4. In some cases the FIFA criteria seemed incapable (weak) of explaining the real performance of the teams. For example the case of some African teams, such as Togo - CAF 2010 events (Reuters, 2010) - and Nigeria. The actual events happened in the team was not compatible with the way their FIFA scores were changing. This is true even in the case of Nigeria. Nigeria has a great drop of position in the FIFA world rankings in 2007. This drop was reportedly due to their 4-1 loss to Ghana in an international friendly match in London. As a friendly match, the loss shouldn't affect the performance score of this team. Of course the number of such cases was few and could be easily found through a simple internet search. We omitted those rare cases from the whole study.

6. CONCLUSION

In this research a framework to monitor team performance was studied. Our model can help to identify whether or not a team shows different performance from what is expected. The differences can be analyzed to determine their cause of happening, and improvement steps can be taken afterwards. We have determined which teams have the same curve base performance. Six groups of performance were discovered. We presented the status of each group and the member teams. Our results showed that even though some teams have close ranks, they may experience different trends. We compared our findings with the outcomes of the FIFA 2010 world cup. The results were interesting and worthy of consideration. The interesting thing about this research is that, according to our study Spain was the only team among the top chart teams that had an increasing trend and had a chance to win the FIFA World Cup 2010 which happened to be the actual outcome of the event. To be effective, team performance status should be regularly monitored with the proposed procedure. Information explored by our proposed approach can be valuable for the sports managers- knowing that their team is in the class of progressive teams or if it is going on a downfall. Also, transition from one performance group to another can have an important significance. Although some experts may not agree with FIFA rankings as a good measure of performance, the aim of this study was to introduce functional data analysis in sport quantitative research. Any kind of functional data from performance measures or other quality parameters can be used by our proposed algorithm. For analyzing the output trends, using the expert's opinion can help to achieve better interpretations. Further research in this area can be done using Statistical Quality Control to monitor the curve-based performance of teams.

Acknowledgements

Thanks to the anonymous reviewers for valuable comments which helped us to correct errors and improve this article.

References

Borrie, A., Jonsson, G. K., & Magnusson, M. S. (2002). Temporal pattern analysis and its applicability in sport: an explanation and exemplar data. Journal of Sports Sciences, 20(10), 845 - 852.

Chen, I., Chen, M.-Y., Jin, C., & Yan, H. (2008). Large-scale Cluster Analysis of Elite Male and Female Swimmers 'Race Patterns. International Journal of Sports Science and Engineering, 123-128.

Chen, I., Homma, H., Jin, C., & Yan, H. (2007). Clustering and display of elite swimmers' race patterns across various comparable criteria at the same time. International Journal of Sports Science and Engineering, 129-136.

David, J. A., Pasteur, R. D., Saif Ahmad, M., & Janning, M. C. (2011). NFL Prediction using Committees of Artificial Neural Networks. Journal of Quantitative Analysis in Sports , 7 (2).

Duan, Z.-H., Liou, L., Shi, T., & Didonato, J. (2003). Application of Singular Value Decomposition and Functional Clustering to Analyzing Gene Expression Profiles of Renal Cell Carcinoma. Proceedings of the 2003 IEEE Bioinformatics Conference.

FIFA. (2011). Retrieved 2 21, 2011, from FIFA/Coca-Cola World Ranking : http://www.fifa.com/worldfootball/ranking/procedure/men.html

Fischer, A., Do, M., Stein, T., Asfour, T., Dillmann, R., & Schwameder, H. (2011). Recognition of Individual Kinematic Patterns during Walking and Running- A comparison of Artificial Neural Networks and Support Vector Machines. International Journal of Computer Science in Sport , 10 (1).

Garcıa-Escudero, L., & Gordaliza, A. (2005). A Proposal for Robust Curve Clustering. Journal of Classification, 22, 185-201.

Goia, A., May, C., & Fusai, G. (2010). Functional clustering and linear regression for peak load forecasting. International Journal of Forecasting, 26, 700–711.

Hastie, T., Tibshiran, R., & Friedman, J. (2009). Elements of statistical learning: Data Mining, Inference, and Prediction. Springer-Verlag.

Heiny, E. L., & Blevins, D. (2011). Predicting the Atlanta Falcons Play-Calling Using Discriminant Analysis. Journal of Quantitative Analysis in Sports, 7 (3).

Jin, S., Ham, B., & Kim, S. (2005). Functional clustering in EEG photic and auditory driving in schizophrenia. International Journal of Psychophysiology, 56, 249– 259.

Lamb, P., Bartlett, R., & Robins, A. (2010). Self-Organising Maps: An Objective Method for Clustering Complex Human Movement. International Journal of Computer Science in Sport, 9 (1), 20-29.

Levernier, W., & Barilla, A. G. (2007). An Analysis of the Home-Field Advantage in Major League Baseball Using Logit Models: Evidence from the 2004 and 2005 Seasons. Journal of Quantitative Analysis in Sports, 3 (1).

McCullagh, J. (2010). Data Mining in Sport: A Neural Network Approach. International Journal of Sports Science and Engineering, 4 (3), 131-138.

Mills, B. M., & Salaga, S. (2011). Using Tree Ensembles to Analyze National Baseball Hall of Fame Voting Patterns: An Application to Discrimination in BBWAA

Voting. Journal of Quantitative Analysis in Sports, 7 (4).
Perl, J., & Memmert, D. (2011). Net-Based Game Analysis by Means of the Software Tool SOCCER. International Journal of Computer Science in Sport, 10 (2).
Pfeiffer, M., & Perl, J. (2006). Analysis of tactical Structures in team handball by means of artificial neural networks. International Journal of Computer Science in Sport, 5 (1).
Puterman, M. L., & Wang, Q. (2011). Optimal Dynamic Clustering Through Relegation and Promotion: How to Design a Competitive Sports League. Journal of Quantitative Analysis in Sports, 7 (2).
Ramsay, J., & Silverman, B. (2005). Functional Data Analysis (second Ed.). Springer.
Reuters (2010-01-09). "African Cup of Nations - NoConfusion over Togo death toll". http://uk.eurosport.yahoo.com/10012010/58/african-cup-nations-noconfusion-togo-death-toll.html. Retrieved 2010-02-01.
Solieman, O. (2006). Data Mining in Sports:A Research Overview.
Song, J., Lee, H., Morris, J., & Kang, S. (2007). Clustering of time-course gene expression data using functional data analysis. Computational Biology and Chemistry, 31, 265–274.

PATH PLANNING OF AN AUTONOMOUS MOBILE ROBOT AS AN ELECTRONIC BALL BOY USING OPTIMAL CONTROL

M. H. Korayem [a, b], I. Khosravipour [a]

[a] Robotic Research Laboratory, School of Mechanical Engineering,
Iran University of Science and Technology, Tehran, Iran

[b] *Corresponding author: hkorayem@iust.ac.ir*

Abstract

Robots are being used for many tasks that are either hazardous or unpleasant for human beings. The potential of robots to assume such tasks depend on their ability to intelligently and efficiently to locate and interact with objects in their environment. One of the applications of mobile robots is capturing and storing of tennis balls. In this paper dynamical formulation as well as experimental studies path planning of non holonomic Wheeled Mobile Manipulator (WMM) which can be used as an electronic ball boy is presented. A Full dynamic model of wheeled mobile base and mounted manipulator is considered with respect to the dynamic of non-holonomic constraint and optimality conditions for carrying maximum payload between start point is derived. Pontryagin's minimum principle which leads to a two-point boundary value problem has been used in this paper. An iterative algorithm is proposed by considering motor restrictions in terms of torque and jerk along the trajectory. Using this algorithm ensures that the resulting trajectory is smooth enough and is optimal in such a way that can be used as an autonomous mobile robot which can gather the balls even in professional tennis games. To verify the accuracy of the algorithm, the proposed approaches are illustrated using simulations and experimental studies of a three-link Scout WMM.

Keywords: Path planning, Mobile robot, Tennis ball collector, Optimal control, Maximum payload

1. INTRODUCTION

Using mobile robots is growing rapidly in human's life and one of the applications that has attracted engineers to itself is modifying them in a way to serve as electronic ball boys, however one of the most important issues regarding manufacturing of these mobile robots is cost reduction and electronic ball boys are no exception. In this article path planning has been regarded to achieve this goal along with improving the quality of product.

In order to evaluate the optimal path the useful method of Pontyagin's minimum principle and optimal control has been used and the resulting two point-boundary value problem has been solved.

The main purpose in this research is to find a path for the robot, in which, it would have minimum jerk while it is trying to capture and store tennis balls it could handle its main task and overcome external obstacles and internal boundaries like non-holonomic constraints, motor's torque and jerk restrictions

It should be mentioned that Korayem and Azimirad (2010) has previously proved that optimal control for path planning is accurate.

Controlling the accuracy and reliability of electronic ball boys would be a critical issue if they are going to be used in professional tennis contests. This evaluation has been presented in this article along with evaluating a new path for these handy sport robots

This paper is organized in the following manner: Section 2 of this paper discusses dynamic modelling of the robot. In section 3, path planning with using optimal control has been discussed. Results of simulation for a three link robot have been provided in section 4 and at last conclusion has been done.

2. DYNAMIC MODLING AND PATH PLANNING OF ROBOT

For initiating the process a non-holonomic wheeled mobile robot manipulator has been considered and its equation of motion has been deduced. This model has covered the governing dynamics of wheels, portable base and the manipulator. The detail about mobile platform which contains a couple of driving wheel and some caster ones has been depicted in Fig 1.

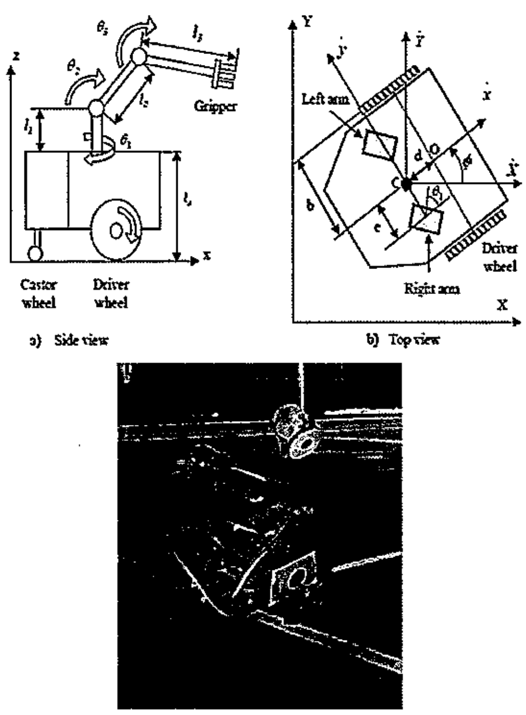


Fig. 1: Picture and schematic of autonomous non-holonomic mechanical mobile manipulator

Three constraints have been considered in this model. First one to be noted is the constraint of moving the base along the axis of symmetry, and the two others are due to the lack of slippage in wheels (driving).

$$\dot{Y}\cos\phi - \dot{X}\sin\phi - d\dot{\phi} = 0 \quad (1)$$

$$\dot{X}\cos\phi + \dot{Y}\sin\phi + b\dot{\phi} = r\dot{\theta}_r \quad (2)$$

$$\dot{X}\cos\phi + \dot{Y}\sin\phi - b\dot{\phi} = r\dot{\theta}_l \quad (3)$$

Considering q as:

$$q = [X \quad Y \quad \phi \quad \theta_r \quad \theta_l \quad \theta_1 \quad \theta_2]^T \quad (4)$$

Using Lagrangian approach for writing the equations of motion for the platform and manipulator the dynamic model of WMM would be derived as following [17]:

$$M_r(q_r)\ddot{q}_r + V_{r1}(q_r,\dot{q}_r) + V_{r2}(q_r,\dot{q}_r,\dot{q}_v) = \tau_r - R_r(q_r,q_v)\ddot{q}_v \quad (5)$$

$$M_{v1}(q_v)\ddot{q}_v + V_{v1}(q_v,\dot{q}_v) + V_{v2}(q_r,q_v,\dot{q}_r,\dot{q}_v) = E_v\tau_v - A^T\lambda - M_{v2}(q_r,q_v)\ddot{q}_v - R_v(q_r,q_v)\ddot{q}_r \quad (6)$$

Where qr and qv represent the two dimensional Lagrangian coordinates of the manipulator and vehicle respectively while Mr and Mv2 represent inertia terms of them.
Vr1 and Vr2 are inertia matrix, velocity dependant terms of manipulator and Coriolis and centrifugal terms respectively. In addition τ_r and τ_v represents the inserted torque of manipulator and vehicle respectively. The effect of the vehicle dynamics on the manipulator has been shown through the Rr matrix. Mass inertia matrix is shown by Mv1 while Vv1 denotes the velocity dependant term of base.
Vv2 represents the Coriolis and centrifugal terms due to the presence of the manipulator.
Ev is a constant matrix and the vector of Lagrangian multipliers associated to kinematic constraints has been named as λ Dynamic effects of the motion made by arm on the whole vehicle has been demonstrated by inertia matrix and is represented by Rv.
It should be noted that this research is for sure more accurate than earlier researches around similar subjects since unlike previous studies that have used the WMM model by only considering the compensation of interaction forces-resulted from the motion of the base- the presented work has faced the full dynamic model of all segments

including platform and manipulator with compensation of interaction forces. In order to do so, it could be seen in equation 9 the dynamic model of non-holonomic WMM has been derived by using Lagrange equation of motion.

$$M\ddot{q} + V(q,\dot{q}) = E\tau - A^T\lambda \qquad (7)$$

By setting state variables as (8), Lagrange multipliers could be eliminated:

$$X = \begin{bmatrix} q^T & v^T \end{bmatrix}^T \qquad (8)$$

$$q = \begin{bmatrix} X & Y & \phi & \theta_r & \theta_l & \theta_1 & \theta_2 \end{bmatrix}^T, \quad v = \begin{bmatrix} \dot{\theta}_r & \dot{\theta}_l & \dot{\theta}_1 & \dot{\theta}_2 \end{bmatrix}^T \qquad (9)$$

In order to evaluate the velocity $\dot{q}$, the null space of Aq should be calculated and as a result matrix S would be derived:

$$\dot{q} = Sv \qquad (10)$$

By defining 7×4 matrix S(q) such that:

$$A(q).S(q) = 0 \qquad (11)$$

$$S = \begin{bmatrix} c(bCos\phi - dSin\phi) & c(bCos\phi + dSin\phi) & 0 & 0 \\ c(bSin\phi + dCos\phi) & c(bSin\phi - dCos\phi) & 0 & 0 \\ c & -c & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad (12)$$

Columns of S in the null space of A(q) matrix. By differentiating Eq. (10):

$$\ddot{q} = S\dot{v} + \dot{S}v \qquad (13)$$

So the kinematic and dynamic equations of non-holonomic mobile manipulator may be represented in the state space form:

$$\dot{x} = \begin{bmatrix} Sv \\ (S^TMS)^{-1}(-S^TM\dot{S}v - S^TV) \end{bmatrix} + \begin{bmatrix} 0 \\ (S^TMS)^{-1} \end{bmatrix}\tau \qquad (14)$$

As mentioned above the optimum path for e-ball boy has been found using optimal control.

The important factor is to determine a possible pathway for all segments in a way that robot would be able to carry maximum payload, knowing the initial and final configuration. Present study has shown that knowing the magnitude of payload the optimum pathway could be found by solving an optimal control problem.

It should be noted that presence of obstacles has not been ignored and limitation of torque and jerk has also been considered.

The overall approach is by using minimum Pontryagin's principle and solving the resulting two point boundary-value problems. The cost function includes velocity and torque terms and related states.

Hamiltonian function is used for optimal control and defined as:

$$H(X,U,\Psi,m_p,t) = L(X,U) + \psi^T(t)F(X,U,m_p) \qquad (18)$$

Principle of minimum Pontryagin implies that following conditions must be satisfied:

$$\dot{X}^*(t) = \frac{\partial H}{\partial \psi}(X^*,U^*,\psi^*,t) \qquad (19)$$

$$\dot{\psi}^*(t) = -\frac{\partial H}{\partial X}(X^*,U^*,\psi^*,t) \qquad (20)$$

$$H(X^*,U^*,\psi^*,m_p,t) \leq H(X^*,\psi^*,\bar{U},m_p,t) \qquad (21)$$

In the above equations (*) represent extremums of X(t), U(t) and ψ(t).

Because of the limitation of control values to an upper and lower boundary the optimal control could be defined as:

$$U = \begin{cases} U^+ & \partial H/\partial U = 0 > U^+ \\ \partial H/\partial U & U^- < U < U^+ \\ U^- & \partial H/\partial U = 0 < U^- \end{cases} \qquad (22)$$

A standard two point boundary-value problem which could be deduced from equations 19 and 20 along with the optimal control of 22 and BCs., could be solved almost simply by approaches like multiple shooting or collocation, note that conditions like accuracy of reaching final goal in specified time and jerk imposed on motors should be satisfied while dealing with a two point boundary-value problem. This accuracy would be evaluated by ε in the following equation:

$$\varepsilon = \left| X(t_f) - X_f \right| \qquad (23)$$

An upper bound for jerk is necessary in order to evade possible errors and the effect of slippage of wheels during tests, along with its effect on limiting the stress imposed on structure.

While running the algorithm of computing MADL an upper boundary for jerk has been evaluated in an indirect method.

In order to conduct the solution approach successfully a two loop algorithm has been considered, in which the first loop is responsible for increasing the payload in each complete loop, and the other one is controlling the intervals. So not only the payload calculation would be accurate but also the approaching rate to the result is reliable.

Since ε is in desirable range and magnitude of jerk meets the limitations the equation 23 is satisfied and payload increases in each iteration of the first loop till the payload value meets its maximum value and passes it. At this moment the jerk of joints and other parts increase so rapidly that equation 23 would no longer be satisfied. Carrying a load more than the maximum payload requires more torque but this would not be a matter of fact cause in each loop of solution, torque constraints would be controlled.

The final point that should be considered is the fact that motors have capacity limitation regarding the torque that is applied on non-redundant joints. This fact is an extra limitation on maximum payload which makes it necessary to find the magnitude of inserted torque on non-redundant joints after solving the two point boundary value problem, using eq.2.

4. RESULTS

For simulation model we set the initial and final point as below:

	A.P. (j.1)	A.P. (j.2)	A.P. (j.3)	A.V. (j.1)	A.V. (j.2)	A.V. (j.3)
Initial point	-pi/2	0	0	0	pi/3	0
Final point	-pi/3	-pi/5	-pi/2	0	0	0
A.= Angular / P. = position/ V. = Velocity/ j. = joint All data are in radian						

Table 1: Initial and final points in test

The optimal path of the end effector and mobile base are shown in Fig. 3. These figures show the angular positions and velocities of wheels and joints also the jerk is presented.

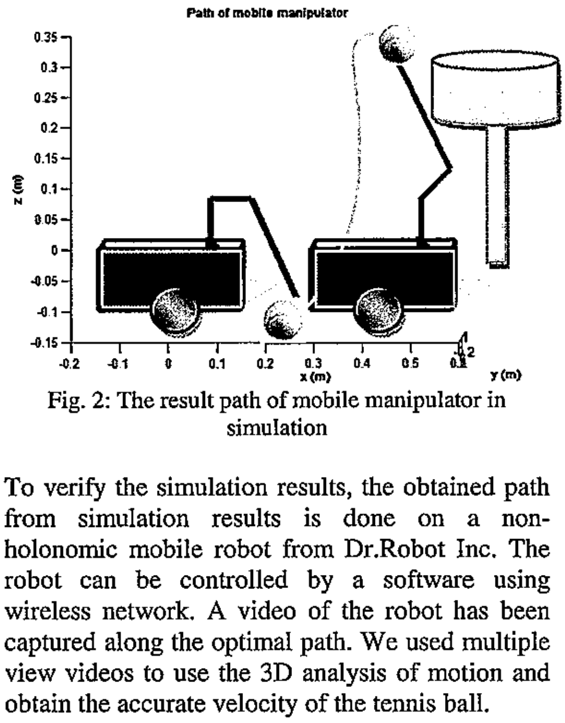


Fig. 2: The result path of mobile manipulator in simulation

To verify the simulation results, the obtained path from simulation results is done on a non-holonomic mobile robot from Dr.Robot Inc. The robot can be controlled by a software using wireless network. A video of the robot has been captured along the optimal path. We used multiple view videos to use the 3D analysis of motion and obtain the accurate velocity of the tennis ball.

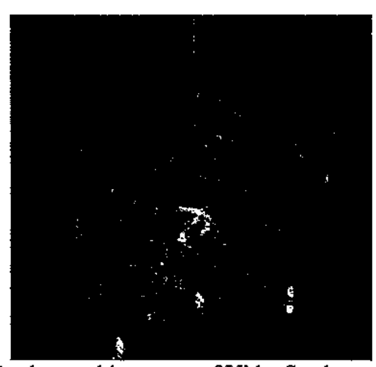


Fig.4: Strobographic output of VideoStrobe software – One view

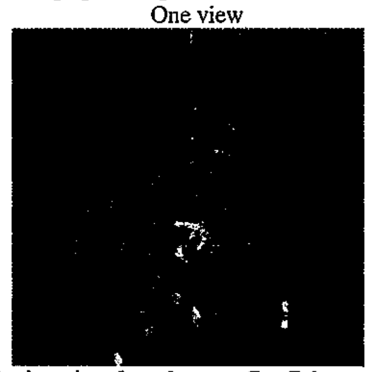
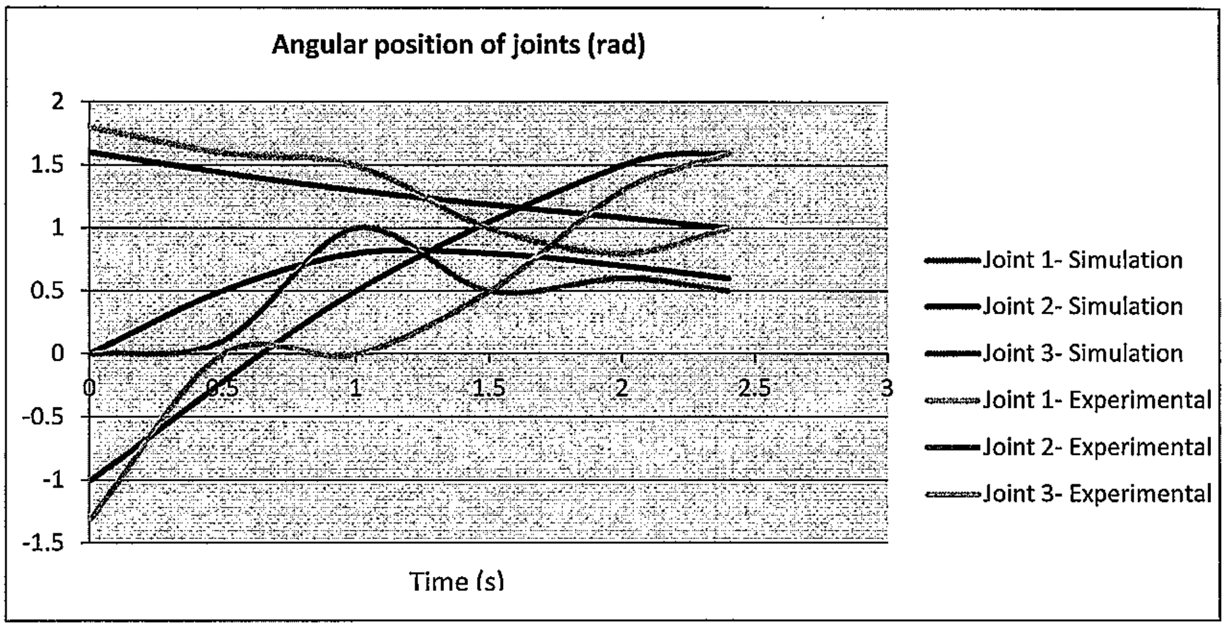
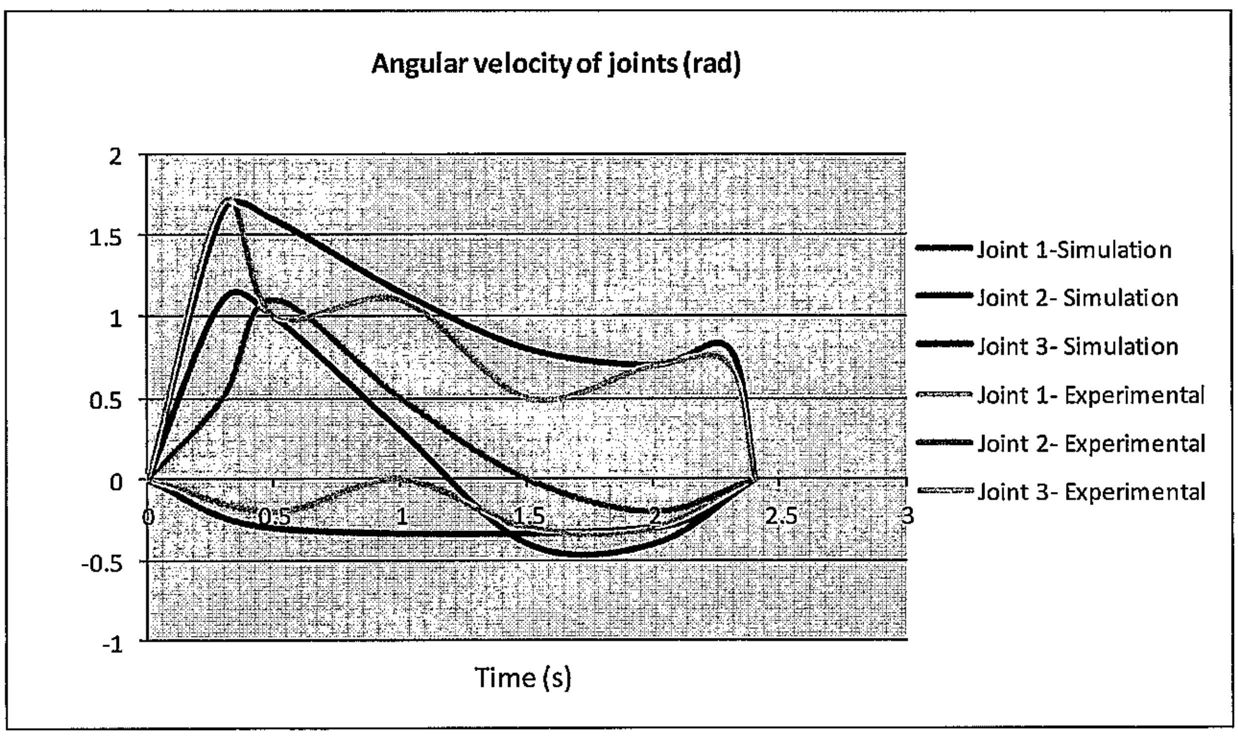


Fig.5: Setting virtual markers on GeoGebra software – one view


Angular position of joints (rad)
2
1.5
1
0.5
0
-0.5
-1
-1.5
0
0.5
1
1.5
2
2.5
3
Time (s)
Joint 1- Simulation
Joint 2- Simulation
Joint 3- Simulation
Joint 1- Experimental
Joint 2- Experimental
Joint 3- Experimental


Angular velocity of joints (rad)
2
1.5
1
0.5
0
-0.5
-1
0
0.5
1
1.5
2
2.5
3
Time (s)
Joint 1-Simulation
Joint 2- Simulation
Joint 3- Simulation
Joint 1- Experimental
Joint 2- Experimental
Joint 3- Experimental

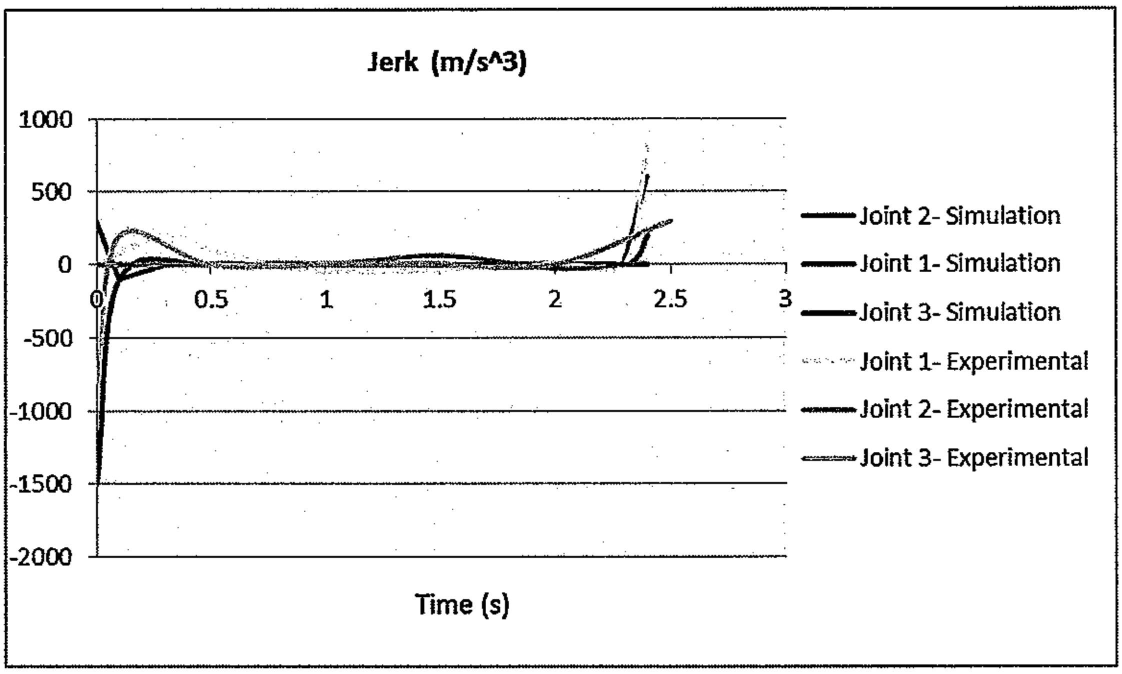


Fig. 3: Simulation and experimental results: (a) Angular position of joints (b) Angular velocity of joints (c) Jerk

Then by use of VideoStrobe software, the analysis of the ball which is in the end effector of robot could be done. By use of this software a strobographic image can be create from a video as shown in Fig 3. Motion analysis can then be carried out with dynamic geometry software such as GeoGebra. The final figure of the analysis can be seen in Fig.4. By use of these softwares, the angular position and velocity and jerk have obtained as Fig. 3. As we can see in Fig. 3, there are few differences between experimental and simulation curves. The most important reason of this difference is the sudden changes in the motors of the robot and the second reason is the low accuracy of the Strobographic method of analysing the video. Setting real marks is a more accurate method of analysing the videos.

5. CONCLUSIONS

The mail target of this paper is path planning of a mobile robot by use of optimal control, so that this mobile robot could be used as electronic ball boy in professional tennis courts. This research has been done on a non holonomic mobile robot with a 3-link manipulator. This path planning has its advantages as below:

First of all, optimization is the main aim of modern life nowadays. So, the more optimized path for a mobile robot-which will be used in tennis courts-, will result in the advantages like being faster, less costly and less energy waster.

Plus, minimum jerk is a perfect parameter to be considered in path planning of robot cause it will result to a smooth motion of manipulators considering the torque boundaries. In addition it prevents the effect of sudden changes in torque on wheels' slippage and motor failure.

References

M. H. Korayem, A. Nikoobin, V. Azimirad, "Maximum load carrying capacity of mobile manipulators: optimal control approach", Robotica, Vol. 27, Pp. 147-159, 2009.

Y.L. Yao, M.H. Korayem, A. Basu, "Maximum allowable load of flexible manipulators for given dynamic trajectory", Robotics and Computer Integrated Manufacturing, v. 10(4), pp. 301-308 -1994Fabianic, D, 1994.

Piazzi A, Visioli A. Global minimum-jerk trajectory planning of robot manipulators. IEEE Transactions on Industrial Electronics -P.145-147 -2000

M. H. Korayem, H. Ghariblu, "Maximum allowable load of mobile manipulator for two given end points of end effectors", The International Journal of Advanced Manufacturing Technology, Vol. 24 (9), Pp. 743-751, 2004.

V. Padois, Y. Fourquet, P. Chiron, "Kinematic and dynamic model-based control of wheeled mobile manipulators: a unified framework for reactive approaches", Robotica, Vol. 25, Pp. 157-173, 2007. L. T. Wang, B. Ravani, "Dynamic load carrying capacity of mechanical manipulators-Part 1", Journal of Dynamic Systems, Measurement and Control, Vol. 110, Pp. 46-52, 1988.

A. Piazzi, A. Visioli, "Global minimum time trajectory planning of mechanical manipulators using internal analysis", International Journal of Control, Vol. 71 (4), Pp. 631-652, 1998.

P. Herman, K. Kozlowski, "Evaluation of manipulator dynamics based on optimal controller", Proceeding of the IEEE Conference on Decision & Control, Pp. 6122-6127, 2006.

R. J. Webster, A. S. Brannon, "The electronic ball boy: a reactive visually guided mobile robot for the tennis court", Robotics and automation, 2054-2059, 2002

H. G. Tanner, K. J. Kyriakopoulos, "Mobile manipulator modelling with Kane's approach", Robotica, Vol. 19, Pp. 675-690, 2001.

Z. Shiller, "Time-energy optimal control of articulated systems with geometric path constraints", Proceedings of the IEEE International Conference on Robotics and Automation, Vol. 4, Pp. 2680-2685, 1994.

M. H. Korayem, V. Azimirad, B. Tabibian, M. Abolhasani, "Analysis and Experimental Study of Non-holonomic Mobile Manipulator in Presence of Obstacles for Moving Boundary Condition", Acta Astronautica, Volume 67, Issues 7-8, 2010

A COMPARISON OF THE PREDICTIVE PERFORMANCE OF DYNAMIC UPDATING METHODS FOR CHESS PLAYER RATINGS

Alec G. Stephenson

CSIRO Mathematics, Informatics and Statistics,
Clayton South, Victoria, Australia.
alec.stephenson@csiro.au

Abstract

This article uses a large dataset to analyze and compare the predictive performance of different dynamic updating methods for rating chess players. The analysis shows that the simpler Elo system is outperformed by both the Glicko and Stephenson systems. The analysis also suggests that the K factors used in the current FIDE (World Chess Federation) implementation of the Elo system are smaller than what would be needed for optimum predictive performance.

Keywords: Chess, Elo, paired comparisons, two-player games

1. INTRODUCTION

Updating systems for rating players (i.e. individuals or teams) in two-player games are fast and surprisingly accurate. The idea is that given games played in time period t, the ratings can be updated using only the information about the status of the system at the end of time period t-1, so that all games before t can be ignored. The ratings can then be used to predict the result of games at time t+1. Comparing the game predictions with the actual results gives a method of evaluating the accuracy of the ratings as an estimate of a player's true skill. There exist more computationally intensive approaches that use the full gaming history via a time decay weighting function (e.g. Sismanis, 2010). These can be more accurate but will not be considered here.

The result of a game is considered to be a value in the interval [0,1]. For chess data, a value of 1 represents a win for white, a value of 0 represents a win for black, and a value of 0.5 represents a draw. The status of the system is typically a small number of features, such as player ratings, player rating standard deviations, and the number of games played. We focus on comparing variations of three basic systems. In increasing order of mathematical complexity, these systems are: the Elo system (Elo, 1978), the Glicko system (Glickman, 1999), and the Stephenson system (Stephenson & Sonas, 2012), which is currently under consideration by FIDE for implementation as the official system for chess player ratings.

The ratings systems considered here derive from statistical models for paired comparisons (e.g. Bradley & Terry, 1952). Preference of one item over another can be related to player preference in two-player games. Draws can be treated as one win and one loss, with each game given a one-half weighting. It is possible to explicitly account for draws using an additional parameter (Davidson, 1970), but this approach does not work well for the dataset considered here. Similar findings for chess data were reported by Joe (1990).

2. METHODS

The dataset we analyse here contains approximately one million games played over the eight year period 1999-2007 by 41,077 chess players. Each record contains the white and black player identifiers, the game result and the month in which the game was played. The dataset was constructed by Jeff Sonas of Chessmetrics and is used with his kind permission. During this period the reporting of individual game results was not required and therefore it contains only a proportion of all games played by FIDE rated players. We also have a dataset of FIDE ratings for 14,118 chess players active at January 1999 which we use to initialize the ratings systems. It may not be an ideal initialization for all systems, but gives a fair method of comparison between them.

We use games from the period 1999-2005 as training data, games from 2006 as validation data, and games from 2007 as test data. Parameter estimation is performed by minimizing the binomial deviance criterion on predictions for the 2006 data, and we evaluate the performance of different systems using the same criterion applied to the unseen 2007 data. For a single game, the binomial deviance criterion is defined by

$$-[S \log(P) + (1-S) \log(1-P)]$$

where S □ {0,0.5,1} is the actual game result and P □ [0,1] is the predicted score. The minimum value is obtained at P=S but predictions of P=0 or P=1 should only be made in cases of 100% certainty otherwise an infinite value could be obtained. For drawn games the minimum value occurs at P=0.5 and is therefore equal to $-\log(0.5) \approx 0.69$. For the overall criterion, we present the mean of this value across all predicted games, multiplied by a scaling factor of 100.

The basic form of the Elo system tracks only the rating R for each player at each time period. After each period, the rating of a player is updated using R □□ $R + K \sum_i (S_i - E_i)$ where the sum is over the games that the player plays within the period, S_i is the actual game result and E_i is the expected game result which is based on the current rating of the player and his or her opponent. The Elo system has one global parameter K which is known as the K factor. The Elo system therefore tracks one system parameter (i.e. the rating) and has one global parameter (i.e. the K factor). In practice the system is often applied by making the K factor dependent on additional information on the player such as the player ratings or number of games played, requiring the use of additional system parameters.

The Glicko and Stephenson systems track both the player rating and the player deviation, which is a measure of the accuracy of the player rating as an estimate of true skill. The mathematical details are more complex and are not given here. The Glicko system has a global parameter c which controls the changes in the deviations through time. In the Stephenson system this role is shared by the global parameters c and h. In addition there is a global neighbourhood parameter λ which shrinks the rating of each player to that of his or her opponents, and a global activity parameter b which gives a small per game bonus irrespective of the result. The b parameter improves predictive performance but also creates rating inflation over time. For chess data this is undesirable and so we do not consider it further.

3. RESULTS

The initialization of ratings is an important issue for all systems. It is useful to distinguish between two forms of initialization: the initialization for players who are already known to exist in the player pool before any updates are performed, and the initialization for players who subsequently enter the system during the updates. For the first case, we use FIDE ratings for 14,118 chess players active at January 1999 as our initial ratings. For the second case, we set the rating of any new player to the value 2200. For the Glicko and Stephenson methods, the initial deviation parameters are set to the value 300 for all players. Another issue in chess is that white typically has a small advantage over black, and this can be modelled using a white advantage parameter γ. It is not important to account for this when constructing the player ratings, but it is important to account for it when predicting subsequent games. We use $\gamma = 30$ for this purpose, which seems roughly optimal across all systems. Table 1 presents the key findings of this article, showing the predictive performance of seven different methods. The seven methods include five different variations of the Elo system, using different methodologies for determining the K Factor. The Elo system is fairly simple, and so several implementations introduce additional complexity by allowing the K factor to depend on additional features. The basic Elo method uses a constant K factor. The methods EloG and EloR use two different K factors. For EloG the K factors are specified according to whether the number of games G played by the player is less than 30, while for EloR they are specified according to

Method	Parameters	Valid (2006)	Test (2007)
Stephenson	$c = h = 9, \lambda = 2$	61.46	62.31
Glicko	$c = 15$	61.54	62.40
EloG ($G < 30$)	$K = 32$ or 26	61.64	62.40
EloR ($R < 2300$)	$K = 32$ or 26	61.63	62.41
EloP ($G < 30, R^* < 2400$)	$K = 30$ or 20 or 15	61.69	62.42
Elo	$K = 27$	61.71	62.47
EloF ($G < 30, R^* < 2400$)	$K = 30$ or 15 or 10	61.96	62.64

Table 1: A comparison of predictive performance of dynamic updating methods for chess player ratings. The Valid and Test columns give the binomial deviance values for predictions on the validation and test data. Details of the different methods are given in the text.

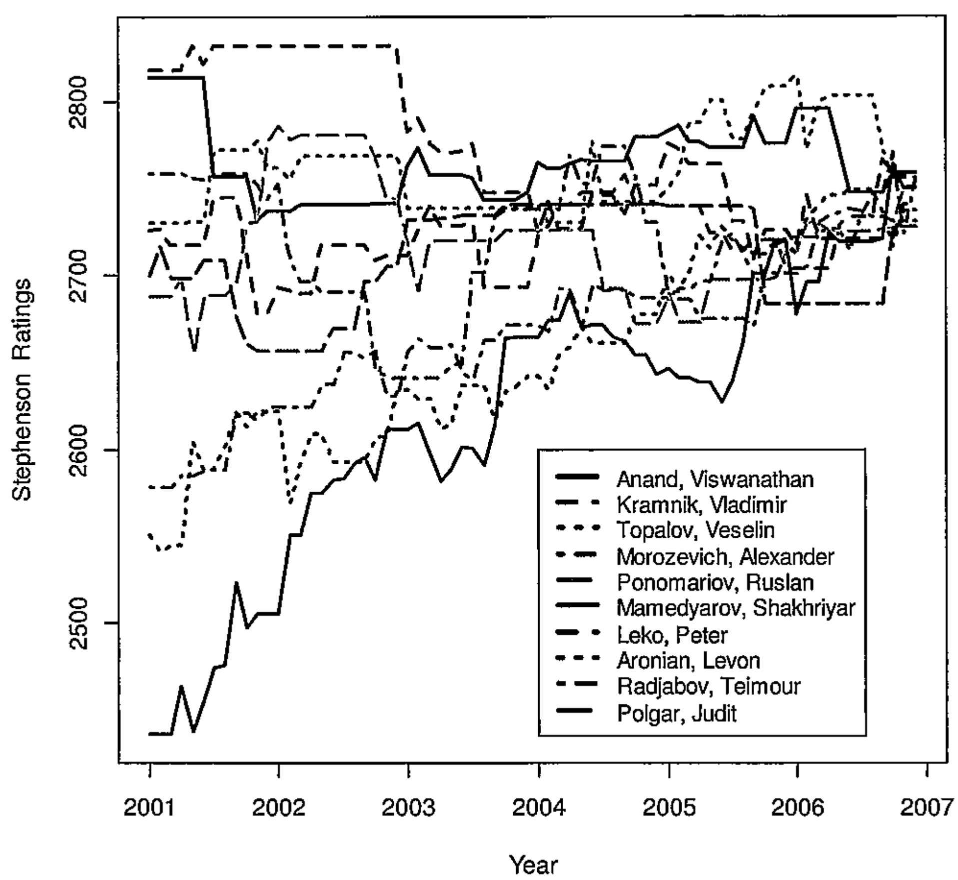


Figure 1: Ratings over time for the `current' (1st Jan 2007) top ten players.

whether the player rating R is less than 2300. Lower K factors are typically associated with more experienced or stronger players, so that their ratings have less tendency to change.

The EloF method applies the FIDE implementation of the K factor. This currently specifies K=30 for players with G < 30 games, K=15 for players with G $\geq$ 30 and whose highest rating ever obtained R^* is less than 2400, and finally K=10 for G $\geq$ 30 and $R^* \geq$ 2400. Although EloF uses exactly the same K factors as FIDE, it does not implement the initialization system of FIDE, which would require knowledge of the type of tournaments that correspond to the games. Despite this, it can still be used to gain some insight into the FIDE ratings implementation. For all methods other than EloF, the parameters have been chosen to be optimal on validation data predictions (i.e. predictions on games in 2006). The EloP method is the same as the EloF method but where K factor values are optimized on the validation data.

The final column of Table 1 shows the predictive accuracy of each method on the unseen test data (i.e. predictions on games in 2007, using data from the period 1999-2006). We see that Stephenson is best, followed by Glicko, then EloG, EloR, EloP, Elo and EloF. The EloF method has the worst predictive performance. The EloP method outperforms EloF because increasing the K factor by 5 for players who have played 30 or more games gives an increase in predictive accuracy.

The top ten players on 1st January 2007 identified by the Stephenson method are shown in Table 2, selecting from the set of players who have played at least 25 games and have played at least once in 2006. The latter condition removes Garry Kasparov. Figure 1 shows the ratings traced over the period 2001-2006 for these same ten players. Note that all rating systems discussed here are relative rating systems, and therefore the mean of the overall ratings is dependent on the method of initialization used in any particular application. The ranking of both the Glicko (not shown) and Stephenson methods are similar, but for Stephenson the absolute ratings are lower. This is a direct consequence of the neighbourhood parameter λ, which draws player's ratings towards their opponents and therefore prevents spread at both the high and low ends. The histogram of the Stephenson ratings (not shown) is slightly more peaked than for Glicko ratings, and acts more like Elo in the upper tail. When $\lambda = 0$, the overall distributions of Glicko and Stephenson ratings are virtually identical, and therefore λ narrows the spread.

	Player	Rating	Deviation	Lag
1	Anand, Viswanathan	2759	65	2
2	Kramnik, Vladimir	2757	61	2
3	Topalov, Veselin	2756	59	2
4	Morozevich, Alexander	2755	60	0
5	Ponomariov, Ruslan	2751	60	1
6	Mamedyarov, Shakhriyar	2750	59	1
7	Leko, Peter	2741	61	1
8	Aronian, Levon	2737	60	1
9	Radjabov, Teimour	2731	61	2
10	Polgar, Judit	2728	65	2

Table 2: The Stephenson ratings and rating deviations for the top ten chess players, 1st January 2007. The lag value represents the number of months since the player last played a game.

The role of the c parameter in Glicko is to increase the rating deviations over time. In Stephenson this role is shared by c and h, and so c is typically lower in Stephenson than the corresponding parameter in Glicko. This feature appears to make little or no difference to the overall distribution of the ratings, but typically improves predictive performance.

4. DISCUSSION

The Elo system has been in existence for more than 50 years. These results suggest that for chess data, rather than attempting to add complexity to the K factor, a better approach for predictive performance is to use systems such as Glicko or Stephenson, which use a rating deviation value to explicitly model the accuracy of the ratings as an estimate of skill. Under these systems, players who have not played many games may have very high or very low ratings with large rating deviation values. It therefore makes sense to consider a rating official only when the player has played some fixed number of games or when the rating deviation decreases below some fixed threshold.

Acknowledgements

We wish to thank Jeff Sonas for providing and archiving the chess dataset.

References

Bradley, R. A. & Terry, M. E. (1952). The rank analysis of incomplete block designs: I, The method of paired comparisons. Biometrika, 39, 324-345.

Davidson, R. R. (1970). On extending the Bradley-Terry model to accommodate ties in paired comparison experiments. J. Am. Statist. Ass., 65, 317-328.

Elo, A. (1978). The Rating of Chessplayers, Past and Present. Arco. ISBN 0-668-04721-6
Glickman, M. E. (1999). Parameter estimation in large dynamic paired comparison experiments. Appl. Statist., 48, 377-394.
Joe, H. (1990). Extended use of paired comparison models, with application to chess rankings. Appl. Statist., 39, 85-93.
Sismanis, Y. (2010). How I won the "Chess Ratings - Elo vs the Rest of the World" competition. arXiv:1012.4571v1
Stephenson, A. G. and Sonas, J. (2012). PlayerRatings: Dynamic Updating Methods For Player Ratings Estimation. R package version 1.0-0.

LOSING THE UNTHINKABLE: AN ANALYSIS OF CHOKING IN ATP TENNIS

Adrian John Schembri[a,b], Elsuida Kondo[a], Anthony Bedford[a], and Stefanie Ristevski[a]

[a] *School of Mathematical and Geospatial Sciences, RMIT University*

[b] *Corresponding author: adrian.schembri@rmit.edu.au*

Abstract

Choking under pressure is a common occurrence in professional sports. In the sporting context, a commonly accepted definition of choking is that it involves the deterioration of an athlete's performance due to real or perceived pressure, despite the athlete's efforts or desires to engage in their optimal performance. There have been limited inquiries into the prevalence of choking during actual sporting events as opposed to experimental environments. The aim of this paper was to investigate the nature and prevalence of choking in ATP tennis, and determine whether choking is more prevalent among certain athletes or stages within tournaments. A database encompassing all best-of-three set matches from 2007 to 2011 was developed. The prevalence of choking among ATP tennis players was explored, as well as the frequency of various types of choking and whether choking was more likely to occur in the second or third sets of a match. Four variations of choking were reviewed; for example, where a player had one the first set, was serving for the match in the second set, but went on to lose the match. Individual players that were most susceptible to choking, as well as player characteristics most clearly associated with choking were identified. In addition, variations in the frequency of choking across each round of a tournament were evaluated.

Keywords: Tennis, Choking, Expectation, Sports Psychology

1. INTRODUCTION

The tendency for players to lose matches despite being in a clear position to win is a phenomenon that has been present in elite sport for centuries (Gentner, 2011). Commonly referred to as choking, this anomaly has been observed in both individual and team sport across both men and women. It has been argued that some individual athletes, teams, or individuals within teams have a culture or predisposition to choke under certain conditions. To illustrate, the Collingwood Football Club in the Australian Football League lost 8 Grand Finals between 1960 and 1981, and earned the nickname 'The Colliwobbles' given their difficulties winning premierships despite being so strong during the home-and-away season. This tendency also extends to individuals sports such as tennis, with Samantha Stosur having a reputation within media circles of choking when playing at major tournaments in Australia. Until recently, even world football champion, Cristiano Ronaldo, was regarded as an under-performer when playing for his national team, Portugal, on the basis of ordinary performances at the 2006 and 2010 Fifa World Cups and the 2008 UEFA Euro Championship. Ronaldo had an exceptional tournament at the recent UEFA Euro 2012, which eased the contention that he had a tendency to underperform when representing the national team.

Choking can be defined as any deterioration in athlete or team performance which is due to real or perceived pressure, despite the athletes' intention or effort to engage in their expected or optimal level of performance (Baumeister, 1984; Baumeister & Showers, 1986). In this regard, pressure is defined as "any factor or combination of factors that increases the importance of performing well on a particular occasion" (Baumeister, 1984, p. 610). Choking may occur at various points of competition, including the beginning, middle periods, and closing stages. Wang (2002) provides an elegant conceptualisation, stating that "after being in a seemingly unassailable position, [the athletes] have choked and subsequently lost in a dramatic fashion" (p. 1).

A challenge associated with the definition of choking is its generality or specificity. Whilst general operationalisations stipulate that choking occurs when an athlete's performance declines relative to expected performance, there is often no mention of the degree that performance must decline before it can be deemed choking. There are also a multitude of reasons underlying why an athlete's performance may begin to suffer throughout a game, such as physical exhaustion (Hornery, Farrow, Iñigo, & Young, 2007), weather conditions, strategic or coaching changes, or injury (Kovacs, 2006).

Choking is a complex phenomenon and is likely to be mediated by several factors. These include the presence of an audience (Wallance et al., 2005); home ground advantage (Koyama & Reade, 2009); length of game play (Gucciardi & Dimmock, 2008); consequences of winning or losing (Gucciardi & Dimmock, 2008); and skill level (Beilock & Carr, 2001). Psychological constructs such as confidence, motivation, mental toughness, and self-efficacy are also likely to have an influential role (Baumeister, Hamilton, & Tice, 1985).

Another challenge of assessing choking in sport is that researchers have had difficulty investigating choking *in vivo*; rather, simulated sporting events have been created. More than two decades ago, research suggested that the mere act of talking about choking was sufficient to significantly increase the prevalence of this phenomenon during a basketball free throw task among 80 male physical education undergraduates (Leith, 1988).

In a more recent study, Gucciardi and Dimmock (2008) assessed the putting performance of 20 experienced golfers, attempting to heighten the risk of loss and perceived pressure by offering financial incentives for better performance. Performance was measured by recording the golfers' mean ball distance from the hole, and comparing all putts against this mean. As hypothesised, the distance of putts varied more from each participant's mean hole distance as pressure increased. The authors acknowledged that it was difficult to simulate the magnitude of pressure that true professionals perceive in high-stake competitions; where the incentives and accolades for winning are much higher than what the researchers could provide (Gucciardi & Dimmock, 2008).

Considerable research had also been conducted on penalty kicks in soccer. Jordet (2009) tested whether professional soccer players with high public-status at the time of the study were likely to choke in high-pressure situations such as penalty shootouts; where audience expectation for these players to succeed was perceived as high. Utilising video and players' statistical information from 366 kicks during three major soccer tournaments, choking was analysed in terms of players' skill level, shot outcomes (goal versus miss), self-regulation behaviours, and current versus future public statuses. The main finding was that significantly fewer goals were scored by current high-status players (65.0%) than by future-status players (88.9%), providing evidence for the impact of perceived pressure on actual performance.

There is limited research on choking during actual professional sporting games. Similarly, little evidence is available to suggest how often choking occurs in elite sport. Given this, it would be advantageous to develop a statistical model to determine the probability that a player has choked, given the likelihood that they would win the match based on the winning position they held earlier in the match. The paucity of research exploring choking instances in professional sport extends to ATP tennis. Whilst tennis players such as Jana Novotna have been outlined in literature as athletes who have been in advantageous situations during matches but have gone on to lose the match, limited research has been conducted (Wang, Marchant, Morris, & Gibbs, 2004). In tennis, the interplay of extraneous performance factors make instances of choking prominent, yet difficult to pinpoint, given the multitude of potential confounding factors. Given the high stakes in tennis matches (players are eliminated from a tournament when they lose a single match), perceived pressure is likely to impact upon player performance, particularly in the latter stages of tournaments.

The aim of the present study was to investigate the nature and prevalence of choking in ATP tennis, and determine whether choking is more prevalent among certain athletes or stages within tournaments. The following research questions were evaluated:

1. What is the prevalence of choking in ATP tennis?
2. Are certain ATP players more (or less) prone to choking during tournaments?
3. Is there a greater prevalence of choking at particular stages of an ATP tournament?
4. Is the prevalence of choking associated with the level of player experience, player quality, or the quality of the opposition?

2. METHODS

2.1 Data Collection

A database of matches played between 2007 and 2011 was developed by collecting ATP match statistics from the Steve G Tennis website (http://www.stevegtennis.com). The exported data incorporated the date of the match, tournament, players, round, player rankings, and match scores. Since only completed matches were of interest, matches such as byes or early retirements were removed. Due to incomplete data in best-of-five-sets matches, only best-of-three set matches were selected for analysis. A total of 10,518 matches were included in the final database. Matches had to meet the following criteria to incorporate potential choking:

1. Best-of-three set matches were required to incorporate all three sets, that is, the player that eventually lost the match must have won at least one set.
2. The deciding set(s) of the match had a final score of 7-6 or 7-5.

These criteria were congruent with Wang's (2002) operationalisation in that the player could have won the match if they had not choked. Out of the sample of 10,518 matches, 3,559 matches were played out to the whole three sets, whilst 656 went to 7-6 or 7-5 in the second and/or third sets.

For those matches classified as potential chokes, OnCourt software (http://www.oncourt.info) was utilised to determine point-by-point information of individual matches. Relevant point-by-point information was used to identify which player was serving at the time, the total number of match points obtained, and the player that won critical points during the match.

2.2 Assessment of Choking

To facilitate the assessment of choking, matches were dichotomised into Type I and Type II scenarios. In a Type I match, the choking player won the first set, but lost the second and third sets, thus match-winning opportunities could have occurred in either the second and/or third sets. In a Type II match, the choking player won the second set, but lost the first and third sets of the match. In Type II matches, it was only necessary to analyse game play in the third set of the match.

The data set was screened for two styles of choking: (i) pure chokes and (ii) tie-break chokes. Pure choke styles were found by first selecting any matches where the deciding set(s) of the match were scored 7-5, then by using OnCourt's point-by-point data to determine if a pure choke style actually occurred. Tie-break choke styles were found by first selecting any matches where the deciding set was scored 7-6, and then by using OnCourt's point-by-point data to determine if a tie-break choke style actually occurred. The definitions and selection criteria of the choking styles are explained below.

2.3 Pure choke

Pure chokes occur when a player has won a set, and leads the second or third sets by a score of 6-5, 5-4, 5-3, 5-2, 5-1, or 5-0, that is, the player is one game away from the match. One or more of the following criteria had to be met for the match to be considered a pure choke:

1. The player was serving for the match on at least one occasion (e.g., If the player was leading 6-2, 5-3);
2. The player was serving for the match and had at least one match point (e.g., If the player was leading 6-2, 5-3 and was winning the next game 40-15);
3. The player was receiving in a match game and had at least one match point (e.g., 6-2, 5-4 and was winning the next game 15-40).

2.4 Tie-break choke

Tie-break chokes occur when the player has won a set, and has at least one opportunity to win the match in a tie break in the second and/or third sets. For example, the score line 6-2, 6-6 would be an example of a tie break choke if the player went on to lose the match. One or more of the following criteria had to be met for the match to be considered a tie-break choke:

1. The player won the first set and played off in a tie break in the second and/or third sets, and had at least one match point (e.g., If the player was leading 6-3, 6-6, and lead 6-4 in the second set tie break);
2. The player won the first set and played off in a tie break in the second and/or third sets, and did not have any match points;
3. The player lost the first set and won the second set, and played off in a tie break in the third set.

3. RESULTS

3.1 Prevalence of Choking in ATP Tennis

Data were collected from 211 professional male tennis players who held an ATP singles ranking and were competing on the ATP circuit. In total, 442 matches were identified as potential scenarios where the player had won the first set and choked in the second and/or third sets. In addition, 214 matches were identified as potential chokes where the player lost the first set, won the second set, yet choked in the third set. After review of point-by-point data using OnCourt, players were found to have choked in 236 matches, that is, they were in what would be regarded as an unassailable position, yet went on to lose the match. Of these, players won the first set and choked in the second or third sets in 168 matches, whilst in 68 matches the player lost the first set, won the second set, and choked in the third and final set. After accounting for the overlap of players competing in multiple matches, 116 players were identified as opponents of chokers and 111 players were identified to have choked on at least one occasion.

Figure 1 displays the frequency of each type of choke that was evaluated in the present analysis.

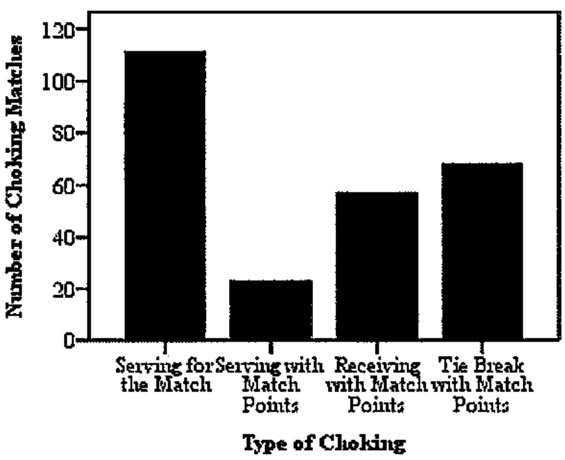


Figure 1. Frequency of the various types of choking in ATP tennis.

The most common type of choke comprised players serving for the match in the second or third sets, yet went on to lose the match. In over 15% of these matches, the player had match points on serve yet went on to lose the match. The second most frequent choking circumstance occurred where the player had match points in a tie break in the second or third set, yet lost the match, with this scenario occurring on over 60 occasions between 2007 and 2011. Slightly less common was a player receiving serve in a match game (e.g., leading 6-4, 5-4) where the player had match points, yet went on to lose the match. Descriptive statistics for each choking category is displayed in Table 1. As shown, 2.2% of all three set matches played between 2007 and 2011 involved choking. Given that the majority of three set tournaments contain 64 players and 63 matches, this equates to approximately one choking incident per tournament. Whilst the number of second and third set chokes was relatively even, quite a high proportion of matches (n = 37) incorporated a player choking in both the second and third sets of a match.

Table 1. Frequency of each type of choke during three set matches between 2007 and 2011.

	Type of Choking				
	1	2	3	4	Total
Frequency	111	23	57	68	236
Proportion of all matches	1%	.2%	.5%	.6%	2.2%
2nd set choke	31	8	11	19	69
3rd set choke	49	10	24	32	115
Both 2nd/3rd set choke	15	5	9	8	37

Note. 1: Serving for the match; 2: Serving for the match with match points; 3: Match points on opposition serve; 4: Match points in tie break.

The most common scenario where a player choked twice within a match was where the player had the opportunity to serve for the match in both the second and third sets. On five occasions, the player had at least one match point when serving for the match in both the second and third sets, yet went on to lose the match. A more detailed analysis of the number of match points players had during best-of-three set matches between 2007 and 2011 is presented in Table 3.

Table 3. Number of match points during matches which eventuated into choking scenarios.

	Set 2		Set 3	
Match Points	S	R	S	R
0	113	96	162	131
1	19	24	29	37
2	5	13	5	17
3	2	3	1	6
4	0	2	1	3
5	1	0	0	4
6	0	2	0	0
Total	40	55	46	84

Note. S: Serving; R: Receiving.

In the majority of matches where players were found to choke, they did not have a match point opportunity. As would be expected, players were more likely to lose match points when receiving serve, with several players losing four or five match point opportunities and then going on to lose the match. On two occasions, a player had six opportunities to win the match when receiving yet went on to lose.

Table 4 displays the players that were found to choke most frequently on the ATP tour between 2007 and 2011.

When comparing all ATP ranked players, John Isner was found to choke most frequently between 2007 and 2011, having choked on seven occasions out of his 171 ATP matches under review. This equated to a choking proportion of 4.09%, the highest proportion of any player. In terms of frequency of choking, Victor Hanescu and Tomas Berdych choked six and five times respectively, followed by several players who choked on four occasions. Of note, Fernando Verdasco and Roger Federer were among the list of players who had choked on four occasions, an intriguing result given Federer was ranked Number 1 in the world for 237 consecutive weeks between 2004 and 2008.

Table 4. Players on the ATP tour who choked most frequently during three set matches between 2007 and 2011.

Player	ATP Matches#	Choking Matches	% of ATP Matches
J. Isner	171	7	4.09%
V. Hanescu	170	6	3.53%
T. Berdych	269	5	1.86%
F. Lopez	207	4	1.93%
F. Verdasco	291	4	1.37%
J. Benneteau	200	4	2.00%
P. Petzschner	121	4	3.31%
R. Federer	274	4	1.46%

#Only matches incorporated in the current analysis are shown here.

In addition, Verdasco was ranked as high as Number 7 in 2009, and made the fourth round in each grand slam during that ATP year. The players who were found to choke least between 2007 and 2011 are displayed in Table 5.

Surprisingly, only five players in total played over 200 matches between 2007 and 2011 and did not choke on at least one occasion. David Ferrer played in the most games between 2007 and 2011 without having choked. Four other players were found to have played in at least 200 matches during the four year period and not have choked, almost all of which could be considered seeded players at Grand Slams during at least one point during the period.

Table 6 displays the players who were most likely to be the opponent of the choking player during the period under review. As shown, Rafael Nadal was the opponent of a player who was deemed to have choked on six occasions, equating to 1.83% of Nadal's matches during this period. Of note, Michael Zverev played in 112 ATP matches between 2007 and 2011, and had an opponent who choked on four occasions, corresponding to a proportion of 3.57%. Other notable players within the top echelon on being the opposition during matches where players choked were Novak Djokovic (Ranked Number 3 in the word for much of the period), Gael Monfils (often criticised for making errors at vital moments of critical matches), and Lleyton Hewitt (often recognised for his mental strength and resilience on-court).

Table 5. Players on the ATP who choked least frequently during three set matches between 2007 and 2011.

Player	ATP Matches	Choking Matches	% of ATP Matches
D. Ferrer	313	0	0.00%
G. Simon	274	0	0.00%
M. Fish	219	0	0.00%
I. Karlovic	208	0	0.00%
S. Wawrinka	206	0	0.00%
J. M del Potro	195	0	0.00%
P-H. Mathieu	181	0	0.00%

Table 6. Players on the ATP tour who players were most likely to choke against during three set matches between 2007 and 2011.

Player	ATP Matches#	Choking Matches	% of ATP Matches
R. Nadal	328	6	1.83%
N. Djokovic	321	5	1.56%
A. Seppi	219	4	1.83%
G. Monfils	225	4	1.78%
J. Isner	171	4	2.34%
J. Benneteau	200	4	2.00%
L. Hewitt	127	4	3.15%
M. Zverev	112	4	3.57%
R. Soderling	240	4	1.67%
S. Querrey	185	4	2.16%

#Only matches incorporated in the current analysis are shown here.

3.2 In-tournament Occurrence of Choking

To evaluate whether choking was associated with the round of matches in ATP tennis, variations in the frequency of choking was assessed in each round of a best-of-three set tournament, refer to Table 7.

Table 7. Occurrence of choking for each Round within best-of-three set ATP tournaments between 2007 and 2011.

	Type of Choking				
Round	1	2	3	4	Total
1	37	12	20	28	97
2	25	9	15	18	67
3	3	0	5	4	12
4	1	1	2	1	5
QF	11	0	3	11	25
SF	8	0	8	4	20
F	2	1	3	2	8

Given that each round contains double the number of matches than each subsequent round, it could be hypothesised that, if the result of chance alone, choking in Round 1 should be double that of Round 2, which should be double that of Round 3 and so on. However, whilst results indicate that choking was most frequent in Round 1 followed by Round 2, the frequency of choking in Quarter Final and Semi Final matches was unusually high, indicating that ATP players were more likely to choke during these latter rounds, over and above what would be expected based on the number of matches played in these rounds. The small number of choking instances in Round 4 is not surprising given that most three-set tournaments contain 64 players in Round 1 and therefore do not have a Round 4, given that the Quarter Finals typically follow Round 3.

3.3 Correlates of Choking in ATP Tennis

Potential correlates of choking were examined to build on past literature which has reported on possible factors that impact on choking, namely experience, player quality and psychological variables such as mental toughness and confidence. Figure 2 displays a scatterplot of the relationship between frequency of choking and ATP matches played.

As shown, a positive relationship exists between choking and match experience, indicating that ATP players who had more experience on the ATP tour were more likely to choke on multiple occasions between 2007 and 2011. A Pearson's correlation was computed to verify this relationship, with a significant result being found, $r(n = 115) = .41, p < .001$. Whilst this may be interpreted to indicate that greater experience increases the likelihood of choking, it is also likely that players who have competed in more ATP matches have had more opportunities to be involved in instances in choking, thus their frequency of choking is often greater than less experienced players.

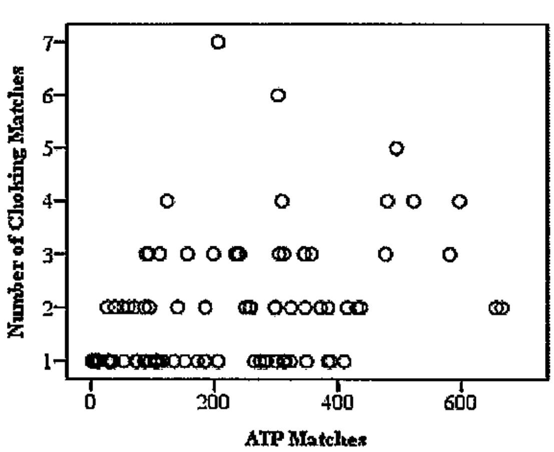


Figure 2. Scatterplot of the relationship between choking and ATP match experience between 2007 and 2011.

Figure 3 displays a scatterplot of the relationship between frequency of choking and player ratings in ATP tennis. Note that in this plot, player ratings are Elo ratings. Again, higher player ratings were associated with more frequent choking on the ATP tour between 2007 and 2011. A Pearson's correlation between player ratings and number of

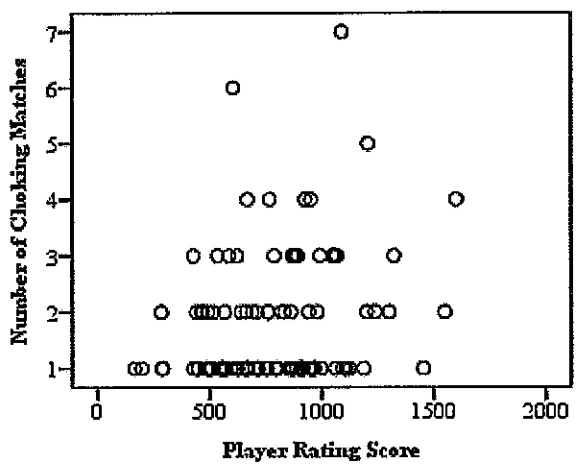


Figure 3. Scatterplot of the relationship between choking and player Elo ratings between 2007 and 2011.

choking matches was found to be significant, $r(n = 115) = .24, p = .017$. One possible interpretation of this finding is that higher ranked players tend to compete in more matches given they qualify for the latter rounds of tournaments, therefore they have more potential to engage in choking given they play in more matches. This may also coincide with more choking during the months of the year where more matches are played. The frequency of choking over time between 2007 and 2011 is displayed in Figure 4.

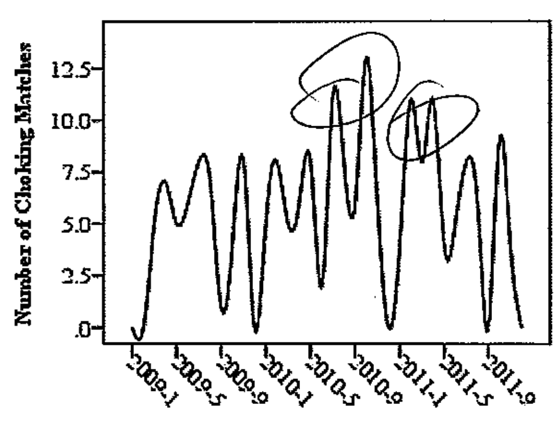


Figure 4. Time series of choking occurrences in ATP tennis between 2007 and 2011.

Choking was found to be low in September of both 2009 and 2011 and during June of 2010, and was lowest during November and December of each year under review, an expected result given very few matches are played during this period. Choking was highest during the second half of 2010 and first half of 2011.

4. DISCUSSION

The present study has presented a novel methodology for the assessment of choking in ATP tennis. Choking was found to be relatively frequent, occurring in 2.2% of best-of-three set ATP matches. The most common type of choking was the circumstance where a player had at least one opportunity to serve for the match, yet went on to lose their service game and subsequently, the match.

Of interest was the finding that a player had match points on serve and went on to lose the match on only 23 occasions between 2007 and 2011. By contrast, on 57 occasions, a player had at least one match point when receiving serve yet went on to lose the match. This finding provides evidence for the likelihood of winning points when on serve, but also the difficulty players have closing out a match when receiving, even if their general play has been superior to their opponent for the duration of the match. The tendency for players to choke in both the second and third sets was almost as high as the frequency of choking in the second or third set alone, indicating that choking in the second set may increase the potential for the player to choke again in the third set.

The results of the current study have provided evidence to suggest that choking occurs both as the result of chance in certain instances, whilst at times, individual differences are related to the likelihood of a choke. Examination of individual players and the tendency to choke revealed that certain players have a greater susceptibility to choking than their peers, with three ATP players found to have choked in over 3% of their matches between 2007 and 2011. The incidence of choking was not found to be limited to lowly ranked ATP players, with well known players such as Tomas Berdych, Fernando Verdasco, and Roger Federer shown to be prominent in relation to choking. Despite this, each of these three players competed in between 250 and 300 ATP matches during the period under review, and therefore their overall choking proportion was found to be less than or equal to 2% of their matches.

The capacity for players to overcome circumstances where defeat appeared inevitable was also evaluated. Rafael Nadal and Novak Djokovic were the top two players to overcome such circumstances, with players found to choke against these two players six and five times respectively between 2007 and 2011. Of note, whilst Roger Federer was found to have choked on four occasions during the period under review, players were found to choke when playing Federer on only 2 occasions. The finding that players who choke are not equally likely to be involved in matches where their opponent chokes provides evidence to suggest that choking does not occur by chance alone. Further evidence that choking is not a chance event was provided by the more frequent occurrence of choking during latter stages of tournaments. Whilst the raw number of choking events was higher in the first two rounds, after accounting for the number of matches in each round, the proportion of choking instances in the quarter and semi finals was higher than in earlier rounds.

The current study has also provided evidence that choking is more likely to be associated with the length of time that a player is on tour, given that matches experience was significantly correlated with choking. The finding that players with higher Elo ratings were also more likely to have multiple choking occurrences provides further evidence for this contention, given that players with higher ratings are more likely to play in more matches given they qualify for the latter stages of tournaments more regularly.

There are several limitations of the current study that should be noted. Firstly, these findings report on only a cross-section of ATP results, given only four years of data was incorporated. Secondly, only best-of-three set matches were included here, and incorporation of five set matches, namely grand slams, will form part of a future analysis. Finally, this analysis focused solely on the ATP tour and did not evaluate choking in the context of women's tennis, which is certainly an area warranting further research.

In order to build on these results, it will be useful to evaluate the relationship between the proportion of matches choked with matches experience and player rating, and evaluate this relationship by means of regression analysis. This will be incorporated in a future analysis of this data. Examination of the proportion of chokes when compared to total wins as opposed to total matches will also build on the present results.

5. CONCLUSION

The present study has presented on the prevalence of choking in ATP tennis. Findings have indicated that choking is most likely to occur in the context of a player serving for the match yet going on to lose the match. Certain players were found to be more susceptible to choking whilst others were more likely to be the opponent during a match where a player choked. Choking was also positively associated with the number of ATP matches a player had competed in, as well as their Elo rating. Choking was most common during the middle periods of the calendar year, a finding that was expected given the proportion of matches is highest during this period, based on the scheduling of Wimbledon, the French Open and the US Open.

6. REFERENCES

Baumeister, R. F. (1984). Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance. Journal of Personality and Social Psychology, 46, 610-620.

Baumeister, R. F., Hamilton, J. C., & Tice, D. M. (1985). Public versus private expectancy of success: Confidence booster or performance pressure? Journal of Personality and Social Psychology, 48, 1447_1457.

Baumeister, R. F. & Showers, C. J. (1986). A review of paradoxical performance effects: Choking under pressure in sports and mental tests. *European Journal of Social Psychology*, 16: 361–383.

Beilock, S. L., & Carr, T. H. (2001). On the fragility of skilled performance: What governs choking under pressure. *Journal of Experimental Psychology, 130*, 701_725.

Gentner, N. (2001). *Choking: What it is and how to avoid it.* Retrieved 8th August, 2012 from http://www.podiumsportsjournal.com/2011/07/23/choking-what-is-it-and-how-to-avoid-it/

Gucciardi, D. F., & Dimmock, J. A. (2008). Choking under pressure in sensorimotor skills: Conscious processing or depleted attentional resources?. *Psychology of Sport and Exercise*, 9, 45–59.

Hornery, D. J., Farrow, D., Iñigo, M., & Young, W. (2007). Fatigue in tennis: Mechanisms of fatigue and effect on performance. *Sports Medicine, 37,* 199-212.

Jordet, G. (2009). When superstars flop: Public status and choking under pressure in international soccer penalty shootouts. *Journal of Applied Sport Psychology, 21,* 125-130.

Kovacs, M. S. (2006). Applied physiology of tennis performance. *British Journal of Sports Medicine, 40,* 381-386.

Koyama, M., & Reade, J. J. (2009). Playing like the home team: An economic investigation into home advantage in football. *International Journal of Sport Finance, 4*(1), 16-41. Retrieved from http://ezproxy.library.usyd.edu.au/login?url=http://search.proquest.com/docview/229402350?accountid=14757

Leith, L. M. (1988). Choking in sports: Are we our own worst enemies? *International Journal of Sport Psychology, 1,* 59-64.

Wang, J. (2002) *Developing and Testing an Integrated Model of Choking in Sport.* PhD thesis, Victoria University.

Wang, J., Marchant, D., Morris, T., & Gibbs, P. (2004). Self-consciousness and trait anxiety as predictors of choking in sport. *Journal of Science and Medicine in Sport, 7,* 174-185.

BETTER APPROXIMATION TO THE DISTRIBUTION OF POINTS PLAYED IN A TENNIS MATCH.

Alan Brown[a, b]

[a] *No affiliation*

[b] Corresponding author: abrown@labyrinth.net.au

Abstract

An estimate of the cumulative distribution of points played in a tennis match based on a straightforward application of the Normal Power (NP) approximation was found to be unsatisfactory. However by partitioning the match endings into disjoint sets it is found that the NP approximation can be applied several times, and the results accumulated to give a better fit. By studying the properties of a final advantage set more closely, an even better overall fit to the cumulative distribution can be achieved. A worked example is given of a best-of-5 final-advantage set match.

Keywords: Tennis, Normal Power approximation, points played

1. INTRODUCTION

The number of points played in a tennis match has a discrete distribution. The moments of this distribution can be calculated using a lattice model with the Markov property and a few other modest assumptions [Brown et al, 2008].

A routine method for recovering an approximation to a distribution is the Normal Power (NP) approximation [Pesonen, 1975]. This method use the first four moments and produces a continuous approximation to the cumulative distribution. The approximation to the frequency distribution can be recovered using differences.

For simplicity we will concentrate on a single example arising from an earlier attempt to find an approximation to the number of points played in a tennis match [Brown et al, 2008], This was a best-of-5 final-advantage set match, where each player has a constant probability of winning a point on serve; 0.77 for player A, and 0.73 for player B, and corresponds to the level attained by two good servers in a professional tournament. The probabilities of winning a standard game on serve are 0.9629 for player A, and 0.9324 for player B.

Since the players are reasonably closely matched the probability of playing the fifth, advantage, set is greater than 30%. The probability of reaching a score of 5-all in this set exceeds 20%, and then the match does not conclude until one of the players attains a lead of 2 games over his opponent. In this example the 98 percentile of the number of points was quoted from a simulation as 576, but the corresponding result of 471 calculated from the NP approximation, was not revealed at that time. A more recent simulation [O'Shaughnessy, 2011] of 50,000 matches put the figure at 575. The aim here is not to argue which simulation result is better, but rather to improve the quality of the calculated result.

Stage 0

statistics	total
Probability	100.00%
Mean	290.34
SD	99.50
Skewness	1.87
Kurtosis	6.21

Table 1. Statistics for a best-of-5 final advantage set; $p_A = 0.77$, $p_B = 0.73$.

The statistics on which the calculation was based is given in Table 1, and the approximation to the cumulative distribution is given in Figure 1. These are labelled as Stage 0 to distinguish it from the developments reported in section 3 of this paper.

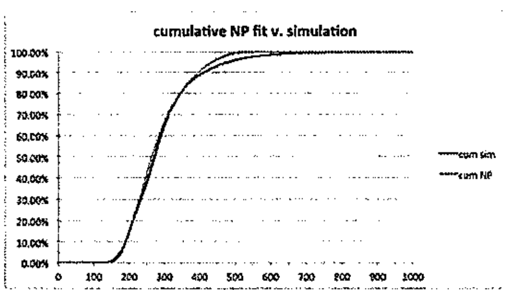


Figure 1. Approximation at stage 0 to the cumulative distribution of points played in a match: $p_A = 0.77$, $p_B = 0.73$.

The approximation to the frequency distribution given in Figure 2 is recovered from the cumulative distribution by differencing.

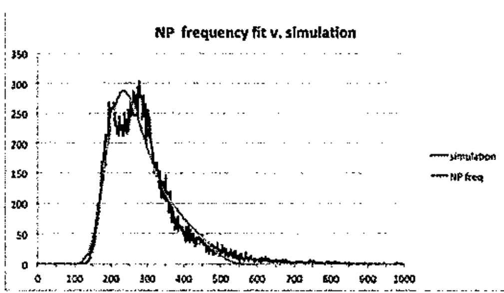


Figure 2. Approximation at stage 0 to the frequency distribution of points played in a match; $p_A = 0.77$, $p_B = 0.73$.

2. The Normal Power approximation

The Normal distribution has been widely studied, and tables of the probabilities for this distribution are readily available. The basic idea of the Normal Power approximation is to use these tables to estimate the tail probabilities of other distributions.

Let X be a random variable with a cumulative distribution F(x), so that

$$\Pr(X \geq x) = F(x)$$

Let μ, σ, γ, κ be the mean, standard deviation, skewness and kurtosis of X. Let Z be a standardised random variable with mean 0 and standard deviation 1, with

$$\Pr(Z \geq z) = \Pr(X \geq x)$$

Denote the cumulative Normal distribution by $\Phi(.)$. Then the approximation can be written as [Pesonen, 1975]

$$F(x) \approx \Phi(y)$$

with

$$z = (x - \mu)/\sigma$$

and

$$y = z - \gamma (z^2 - 1)/6 - \kappa (z^3 - 3z)/24$$
$$+ \gamma^2 (4z^3 - 7z)/36 + \ldots \quad (1)$$

It is well known that the distribution of a random variable is not uniquely determined by its moments. In order for this approximation to be successful, the target distribution must be similar in some sense to the Normal distribution. The Normal distribution is continuous, whilst the distribution of points played in a match is discrete. So the best we can hope to achieve is a good approximation. An exact fit cannot be achieved with the NP approximation.

The NP approximation has several weaknesses that are of some concern:

(a) The NP approximation preserves the unimodal property of the Normal distribution and so fails to replicate the multimodal property of the distribution of points played.

(b) The NP approximation produces a distribution with Normal tails, and fails to fit a distribution with exponential tails. Thus is inappropriate to use in conjunction with the statistics for a tennis match where an advantage set might be played. The long tail of the distribution of the points in an advantage set arises when two good servers are opposed to each other. There was a spectacular illustration of this effect at Wimbledon in 2010, when the match between John Isner and Nicholas Mahut lasted for 3 days. The fifth set ended with a game score of 70-68.

(c) NP approximation is based on an asymptotic expansion. It can become numerically unstable in some circumstances, such as when γ is close to zero and κ is positive.

If we differentiate equation (1) we find that

$$\frac{dy}{dz} = 1 - \gamma/6 - \kappa (z^2 - 1)/8$$
$$+ \gamma^2 (12z^2 - 7)/36 + \ldots \quad (2)$$

When γ is close to zero and κ is positive, equation (2) reduces to

$$\frac{dy}{dz} \approx 1 - \kappa\,(z^2 - 1)/8 \qquad (2')$$

and so, in this case, there is a turning point near

$$z = (8/\kappa)^{1/2} \approx 4 \text{ when } \kappa \approx 0.5$$

Such a turning point cannot occur for a cumulative distribution. A heavy-handed way of avoiding this type of difficulty in the NP approximation is to set

$F(x) = 1$ whenever $z > 4$, and

$F(x) = 0$ whenever $z < -4$.

Fortunately this type of adjustment was not required in the calculations that arise in our example.

3. Steps towards improving the approximation

3.1 Stage 1

We begin by attacking the problem that the distribution of the points played in a match is multimodal. We do this by breaking the ending of the match into three disjoint parts; matches ending in 3 sets, matches ending in 4 sets, and matches ending in 5 sets. Denote by X the number of points played in the match. Then

$$\Pr(X \geq x) = w_3 \Pr(X \geq x \mid \text{match ends in 3 sets})$$

$$+ w_4 \Pr(X \geq x \mid \text{match ends in 4 sets})$$

$$+ w_5 \Pr(X \geq x \mid \text{match ends in 5 sets}) \quad (3)$$

where, for j = 3, 4 or 5, the weights

w_j = probability of match ending in j sets.

The NP approximation can be applied using the moments for each of the disjoint match endings. The three cumulative distributions can then be weighted by the appropriate probabilities, and accumulated as in (3) to give a new approximation to the overall distribution of the number of points played in a match. Working with cumulative distributions rather than frequency distributions avoids the computational problem of convolutions.

The statistics on which this calculation is based are given in Table 2.

Stage 1

statistics	3 set	4 set	5 set	total
Probability	28.96%	37.40%	33.65%	100.00%
Mean	200.46	268.83	391.61	290.34
SD	23.75	27.72	101.79	99.50
Skewness	0.01	-0.07	2.01	1.87
Kurtosis	-0.35	-0.21	6.27	6.21

Table 2. Statistics for various set endings in a best-of-5 final advantage set; $p_A = 0.77$, $p_B = 0.73$.

The approximation to the cumulative distribution is shown in Figure 3. We see that we have achieved better fit, but there is room for improvement in the fifth set.

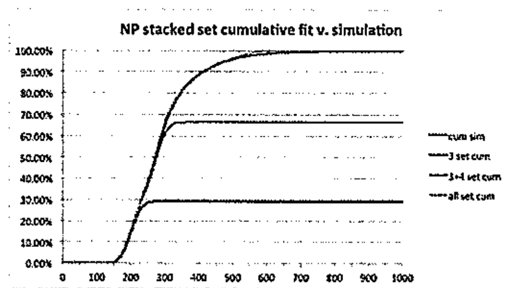


Figure 3. Stacked view of the cumulative distribution of points played in a match: $p_A = 0.77$, $p_B = 0.73$.

3.2 Stage 2

We attack the problem in the fifth set by a further partition the ending of the match in this set into 2 disjoint parts; endings before 5-all, and endings after 5-all. Then

$$w_5 \Pr(X \geq x \mid \text{match ends in 5 sets})$$

$$= w_{5h} \Pr(X \geq x \mid \text{match end before 5-all in } 5^{th} \text{ set})$$

$$+ w_{5t} \Pr(X \geq x \mid \text{match end after 5-all in } 5^{th} \text{ set}) \quad (4)$$

where the weights

w_5, w_{5h}, w_{5t}, = probability of the respective match endings.

The statistics on which this calculation is based are given in Table 3.

Stage 2

statistics	3 set	4 set	5 head	5 tail
Probability	28.96%	37.40%	12.63%	21.01%
Mean	200.46	268.83	322.00	433.46
SD	23.75	27.72	28.39	106.96
Skewness	0.01	-0.07	-0.08	1.80
Kurtosis	-0.35	-0.21	-0.17	5.20

Table 3. More statistics for various set endings in a best-of-5 final advantage set; $p_A = 0.77$, $p_B = 0.73$.

Figure 4 shows that again we have obtained a better fit, but there remains room for improvement after 5-all in the fifth set.

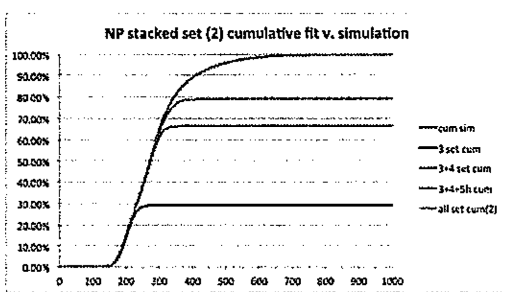


Figure 4. Stacked view of the cumulative distribution of points played in a match after partitioning the endings in the fifth set: $p_A = 0.77$, $p_B = 0.73$.

3.3 Stage 3

The improvement in fit in the first two stages was due to the use of more information, not a change in method. In an attempt to improve the approximation in the closing stages of the fifth set we try something different.

The probability distribution for the number of games in an advantage set is discrete. However the shape of the tail after 10 games is closely related to that of the exponential distribution. After the score of 5-all has been reached the set does not end until one player has achieved a lead of two games over the opponent. Using the condition that the probability of each server winning his game remains constant, it is easy to show that the probability of progressing from one level game score to the next is a constant less than 1. It follows that all the probabilities of future level game scores are geometrically decreasing after five games all. The number of points played in each pair of games between level games scores can vary about a mean, but the overall shape of the distribution of the number of points played after 5-all follows the distribution of the number of games played.

The shifted exponential distribution is a continuous analogue of the geometric distribution, and can be fitted using just two parameters, the mean and the standard deviation.

$$F(x) \approx 1-e^{-(z+1))} \qquad \text{if } z > -1,$$

$$F(x) \approx 0 \qquad \text{if } z \leq -1, \qquad (5)$$

The skewness and kurtosis of the shifted exponential distribution are 2 and 6 respectively, and this is in reasonable agreement with the statistics for the tail of the distribution of points played in an advantage set.

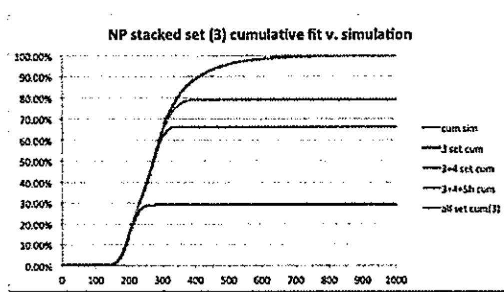


Figure 5. Stacked view of the cumulative distribution of points played in a match after partitioning the endings in the fifth set and using an exponential tail: $p_A = 0.77$, $p_B = 0.73$.

Although using the exponential distribution after 5-all in fifth set seems a likely prospect, checking Figure 5 reveals a problem with "fitting the tail on the donkey", i.e. between the 70th and 80th percentile of the cumulative distribution.

3.4 Stage 4

When we study the fit of the NP approximation to the cumulative exponential distribution, as in Figure 6, we note that there is the good fit in the middle of the distribution, but the poor fit in both tails. We also note that there are two crossing points.

We propose to use a mixture of the NP approximation and the exponential distribution for the distribution of the points played after 5-all in the fifth set, so we must make a choice between the crossing points. Testing quickly shows the crossing point greater than the mean is the better choice.

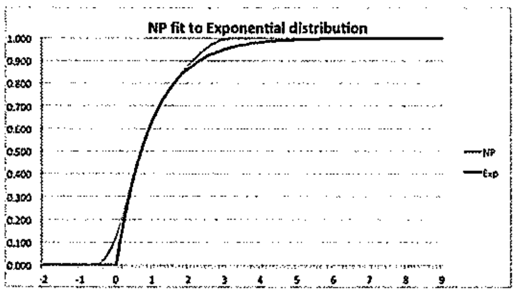


Figure 6. NP approximation to the cumulative distribution of the standard exponential distribution

Our mixture model is using statistics for matches ending after 5-all in 5 fifth set is

$$F(x) \approx \Phi(y) \qquad \text{if } z \le 0, \text{ and}$$

$$F(x) \approx \min(\Phi(y), 1-e^{-(z+1)}) \quad \text{if } z > 0. \qquad (6)$$

This model is then weighted by the probability of this type of ending occurring in a match.

The check of Figure 7 shows that we have achieved a better approximation for all values of x. At the changeover point in the mixture there is a discontinuity in the slope of the approximation to the cumulative distribution. Can you spot the jump in the fitted frequencies between 496 and 497 points in Figure 7?

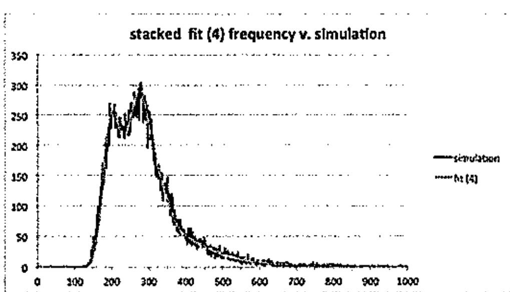


Figure 7. Fit of the frequency distribution of points played in a match after partitioning the endings in the fifth set and using a mixture in the tail: $p_A = 0.77$, $p_B = 0.73$.

4. Discussion

The technique of partitioning the match endings into a few disjoint sets results in an improved approximation for much of the distribution of the points played in a match whilst still making it practical to continue using the NP approximation. The refinement of using a mixture of the NP approximation and the exponential distribution for the tail of endings in the fifth advantage set reduces the remaining errors to an acceptable level.

The quality of the improvement in the approximation at the upper percentiles of the cumulative distribution is shown in Table 4.

Percentile \ stage	0	1	2	3	4	simulation
75%	327	322	326	332	326	328
98%	471	542	558	578	578	575
99%	491	558	597	652	652	649

Table 4. Various estimates of the 75th, 98th and 99th percentiles of the points played in match: $p_A = 0.77$, $p_B = 0.73$.

Although we have discussed only a single example, the methods used here can be applied more widely. In particular, the methods can be adapted to produce an acceptable approximation of the distribution of the remaining number of points to be played in a match in progress. Somewhat perversely the quality of the approximation gets worse as the match progresses. To see how this arises, consider the situation at match point; "It may all be over in two minutes, or two hours."

Acknowledgements

The simulation results for the number of points played in a tennis match provided by Darren O'Shaughnessy and Ian Lisle are gratefully acknowledged.

References

Brown A., Barnett T., Pollard Geoff, Lisle I., and Pollard Graham (2008). The characteristics of various men's tennis doubles scoring systems, Proceedings of the Ninth Australasian Conference on Mathematics and Computers in Sport

Brown A., Barnett T. and Pollard G. (2008). A recursion method for evaluating the moments of a nested scoring system, Proceedings of the Ninth Australasian Conference on Mathematics and Computers in Sport.

Pesonen E. (1975). NP-technique As A Tool In Decision Making. ASTIN Bulletin. 8(3): 359-363.

O'Shaughnessy, D. (2012). Private communication.

DEVELOPING A TENNIS CALCULATOR TO TEACH PROBABILITY AND STATISTICS

Tristan Barnett [a,b]

[a] *University of South Australia*
[b] *Corresponding author: strategicgames@hotmail.com*

Abstract

The longest professional tennis match, in terms of both time and total games occurred at the 1st round of Wimbledon 2010 between John Isner and Nicholas Mahut. It lasted 183 games, required 11 hours and 5 minutes of playing time, with Isner winning 70-68 in the advantage final set. Even with the introduction of a tiebreak set at Wimbledon in 1971 long matches still occur and records of long matches can still be broken. Was this long match predictable and what are the chances of this record being broken in the future? This talk will provide insights to these questions by formulating a mathematical model that provides information such as chances of players winning the match, reaching the advantage final set and reaching 68-68 all in an advantage final set. Hence the mathematics of tennis is concerned with the chances of players winning the match (who is likely to win?) and match duration (when is the match going to finish?). These calculations are required prior to the start of the match, but also for the match in progress. For example, what are the chances of player A winning the match in 4 sets from 1 set all, 3 games all, 30-15 and player A serving? Whilst the mathematics of tennis could be of interest to tennis organizations, commentators, players, coaches and spectators; it could also be applied to teaching by using the well-defined scoring structure of tennis to teach concepts to students in probability and statistics. Such concepts include summing an infinite series, Binomial theorem, backward recursion, forward recursion, generating functions, Markov chain theory and distribution theory. The mathematics of tennis applied to teaching also allows students to build their own tennis calculator using spreadsheets, which in turn could assist in the understanding of probability and statistical concepts, and familiarize students with using spreadsheet software such as Excel. Thus, it is shown in this talk how recursive formulas developed in spreadsheets are used to obtain from any score line within the match; the chances of winning, distributions on the length, and typical parameters of distribution such as the mean with an associated standard deviation.

Keywords: tennis scoring systems, teaching, recursive formulas, spreadsheets

References

Barnett T and Clarke SR (2005). Combining player statistics to predict outcomes of tennis matches. *IMA Journal of Management Mathematics*. 16 (2), 113-120.

Barnett T (2006). Mathematical modelling in hierarchical games with specific reference to tennis, PhD thesis, Swinburne University of Technology, Melbourne .

Barnett T, O'Shaughnessy D and Bedford A (2011). Predicting a tennis match in progress for sports multimedia. *OR Insight* 24(3), 190-204.

Brown A, Barnett T and Pollard G (2008). A recursion method for evaluating the moments of a nested scoring system. In proceedings of the Ninth Australian Conference on Mathematics and Computers in Sport.

Noubary R (2010). Teaching mathematics and statistics using tennis, In J.A. Gallian (Editor), *Mathematics and Sports*, Mathematical Association of America, 43, 227-240.

THE EFFECT OF EASE OF WIN ON PERFORMANCE IN MEN'S TENNIS

Elsuida Kondo, Anthony Bedford[a], Alison Dias and Adrian J Schembri

School of Mathematical and Geospatial Sciences
RMIT University, Melbourne, Australia
[a]Corresponding author: Anthony.bedford@rmit.edu.au

Abstract

Much research is focused on prediction success in tennis matches. The focus on game, set and point estimates consider factors such as winning on serve, and in some instances, the dependent previous results of the two players. However, no such research has looked at the potential influence of the length of the previous match. In this research we develop a metric, Ease of Win (EoW), used to examine whether a player's performance in the previous match of a tournament influenced the subsequent round. This is then compared relatively to the opponent's EoW. This approach is not only aimed towards its incorporation into the SPARKS model, a rating system used here to control for players' relative difference in strength (Bedford & Clarke, 2000), but to demonstrate the potential impact previous matches have on player performance.

EoW for a player ranges between 0 and 1 and is dependent on the number of games and sets won in a match, with players being granted more credit for games won in the later sets. If a player has had an easy win in the previous round, they are rewarded a high EoW score.

In this study, the EoW approach is considered within-tournament using ATP data over nine years (2003 – 2011, n= 13,228). Nested EoW, which is a running average of the EoW across the rounds, was controlled for. Overall results suggest a systematic effect on the win percentages, after accounting for expected form (SPARKS residuals). When we consider EoW independently of expected form, the pre-match predicted favourites with an easier previous match win 5.71% more matches than those with a tough previous match. The results obtained strongly reinforces the hypothesis that pre-match EoW influences outcomes in the next round as there is a systematically higher likelihood of winning when there is a positive EoW. Embedding EoW into the ratings model (SPARKS) improves the models predictive results by 2.12% (71.05% vs. 68.92%). The consistent positive effect of EoW indicates it is an important variable to consider in the performance, and potential performance, of tennis players within tournaments.

Keywords: ease of win, prediction, SPARKS, performance

1. INTRODUCTION

Tennis is a highly participative world class competitive and recreational sport, making it one of the most popular sports in the world. Unlike other major sports, tennis is played on a wide variety of surfaces. Matches are played in the best of three or five sets, depending on the type of tournament. For example the four major Grand Slams (Australian Open, French Open, Wimbledon and US open) are played on different surfaces (hard court, clay and grass) and are all played as the best of five sets. These factors may often influence a players' performance, match intensity and the length of the game. Generally a tennis match can last longer than an hour but in the past there have been occasions where players have had to play longer than five hours.

Nevertheless associated with it popularity as a sport, it is a growing source of income emerging from associated markets, especially wagering. Therefore it comes as no surprise that mathematical sport modelling is widely used for prediction not only in tennis but in many other sport areas. Although applying mathematical models to tennis is not a new concept, there are several variables involved in tennis that perhaps need to be taken into account for more accurate match predictions.

From the existing literature it is clear that there has been much focus on game, set and point estimates, which consider factors such as winning on serve, and in some instances, the dependent previous results of the two players. Also as mentioned earlier tennis is played on a wide variety of surfaces, with different ball types, and matches are played as the best of three or five sets. Regulations to the scoring system, length of match, playing surface, and ball type have been reported to affect the physical and physiological strains of tennis match play (Smekal, 2001, Chandler, 1995, O'Donoghue, 2001, Ferrauti, 2001 and Girard, 2004). In theory these variables could help identify the player who is more likely to dominate the match.

However, no research has looked at the potential influence of the length of the previous match. In this research we develop a metric, Ease of Win (EoW), used to examine whether a player's performance in the previous match of a tournament influenced the subsequent round. This is then compared relatively to the opponent's EoW. This approach is not only aimed towards its incorporation into the SPARKS model, a rating system used here to control for players' relative difference in strength (Bedford & Clarke, 2000), but to demonstrate the potential impact previous matches have on player performance. EoW for a player ranges between 0.2 and 1 and is dependent on the number of games and sets won in a match, with players being granted more credit for games won in the later sets. If a player has had an easy win in the previous round, they are rewarded a high EoW score.

2. METHOD

The Ease of Win method considers player performance in a match and its potential influence on the subsequent round. Here the length of a match is considered to be a performance indicator. The EoW as the name suggests is calculated only for the winner of a match. The formula considered takes into account the sets and games won in a match accounting for the opponents' performance. A player is granted more credit for games won in the later sets as at this stage he has far less resources left. This EoW ranges between 0.2 and 1 with 1 indicating a very easy win. An EoW score of 1 would thus correspond to a win with a score of 6-0 6-0 in a match.

The formula used to determine the Ease of Win is as follows, for a 3 set maximum match:

$$EoW = \frac{1}{6}G_1 + \frac{2}{6}G_2 + \frac{3}{6}G_3 \qquad (1)$$

where, $G = \frac{No.of\ games\ won\ in\ a\ set\ i}{n_i}$

$$G_i = 1\ if\ n_i = 0$$

A multiplier with increasing weights in each set is used.

In cases where the player wins the match in the first two sets, the G_i for set three is given as 1. i.e. they are awarded full credit for the last set. (1 equates to a score of 6-0 for the set)

Similarly, for a five set maximum match the formula is as follows:

$$EoW = \frac{1}{15}G_1 + \frac{2}{15}G_2 + \frac{3}{15}G_3 + \frac{4}{15}G_4 + \frac{5}{15}G_5 \qquad (2)$$

Here in cases where the player wins the match in 3 sets, the 4th and 5th set is given a G_i of 1 and in cases where he wins in 4 sets he will be awarded full credit for 5th set.

Case study

Consider the 2011 finals of the Barcelona tournament played between Rafael Nadal and David Ferrer. This was a maximum three set tournament and Rafael Nadal won with a score of 6-2 6-4. Since he won the match in two sets his ease of win from (1) is calculated as:

$$EoW = \frac{1}{6}\cdot\frac{6}{8} + \frac{2}{6}\cdot\frac{6}{10} + \frac{3}{6}\cdot 1 = 0.825$$

Thus 0.825 indicates a relatively easy win (as we see from the score). Similarly, consider the 2008 Wimbledon finals held in London, played between Rafael Nadal and Roger Federer. Rafael Nadal won the match with a final score of 6-4 6-4 6-7 6-7 9-7. The score definitely indicates that he had a tougher win than in the previous example.

The EoW in this case is calculated to be:

$$EoW = \frac{1}{15}\cdot\frac{6}{10} + \frac{2}{15}\cdot\frac{6}{10} + \frac{3}{15}\cdot\frac{6}{13} + \frac{4}{15}\cdot\frac{6}{13} + \frac{5}{15}\cdot\frac{9}{16} = 0.523.$$

Thus, this EoW score of 0.523 indicates a difficult win.
Alternatively one of the most difficult wins recorded so far in our dataset (detailed later) is 7-6 4-6 7-6 with an EoW score of 0.48. This is because the winner exhausted a lot of his resources in the second set of which he acquired a game rate of 0 in this set. On the other hand, as mentioned earlier the easiest win possible would be a score of 6-0 6-0 which would give an EoW score of 1.
A lot of matches are recorded where a player wins as a result of his opponent retiring on account of injury as well as where matches a player wins as a result of a 'bye' on account of his opponent dropping out for various reasons. These wins are given a fixed EoW score of 1 as their win did not require exhaustion of their available resources and is considered to be very easy. The actual EoW calculated for a match is then used to determine potential influence on players' performance in the next round. Thus, pre-match EoW score in Round 2 for Player 1 and Player 2 would be their EoW score calculated in Round 1. This score is used to help determine impact on player performance in Round 2. In this study, the EoW approach is considered within-tournament. The pre-match EoW in each round is computed looking at results of previous rounds in the tournament.
Consider the following study; Novak Djokovic in the 2011 Madrid tournament.
Round 1 – Djokovic 'won' by a bye thus giving him an EoW score of 1.Thus his pre-match score for Round 2 is 1. In Round 2 Djokovic defeated Kevin Anderson (6-3 6-4) giving him an EoW score of 0.8111.
Thus his pre-match EoW score for Round 3 is a running average of EoW score across round 1 and 2 with more weightage/credit granted to round 2. This is calculated as;

$$\frac{1}{3}EoW_1 + \frac{2}{3}EoW_2 = 0.887$$

Similarly, the scores are computed for subsequent rounds. Thus nested EoW can be defined as a running average of the EoW across the round which is controlled for with more credit granted to the latest round.
The data used for the analysis here includes all Men's ATP Tournaments over the past nine years (2003 – 2011). Round 1 of every tournament is used to compute pre match score for Round 2. However, it is not included in final analysis as Round 1 itself does not have a pre match score. Thus performance pre round 1 cannot be investigated. Thus, a total of 13,228 matches was examined from round two onwards.
This pre-match EoW score is the basis of our analysis in this paper. Our ratings system (SPARKS) as we know helps determine our favourite for every match. The pre-match EoW score for these favourites and their opponents are calculated and then further analysed.

3. RESULTS & DISCUSSION

All results obtained in this paper are in terms of our favourites determined by SPARKS. For every match from 2007 – 2011, we obtain a pre-match EoW score for the favourite and his opponent as discussed earlier. The difference between these scores is then calculated to obtain the Pre-match EoW margin for each match. A positive margin is thus interpreted as in 'our favourite has had an easier win in his previous round match as compared to the win of his opponent.' Similarly a negative margin is interpreted as 'our favourite has had a more difficult win in the previous round as compared to the win of his opponent'.

The range of these margins was approximately 0.6 to - 0.6 with 0.6 indicating a very easy previous win (favourite) and (-0.6) indicating a very difficult previous win (favourite) compared to their opponents. Based on these margins the matches were divided into different positive bands 0.01 – 0.1, 0.11 – 0.2, 0.21 – 0.3,0.31-0.4, >0.4 etc. which could be more easily identified as slightly easy, fairly easy, moderately easy, significantly easy and very easy previous round matches relative to their opponent. Similar bands were created for the negative margins as well which could be identified as slightly tough, fairly tough, moderately tough etc. as compared to their opponent.

Further, using SPARKS the pre match ratings of the players in a match were compared and the difference taken to obtain a pre-match SPARKS margin in terms of our favourite. These matches were divided into different categories as well. The matches where both players had an equal rating was not taken into account for analysis as there was no favourite. For the rest of the matches the

categories for the margin were as follows; 0.1 – 320, 321 – 640, 641 – 960, 961 – 1280, 1281 – 1600 and >1600.
In this case the band 0.1 -320 ratings margin refers to very similar quality of players whereas >1600 would refer to lopsided matches where the favourite has a very high rating and his opponent a very low rating. e.g. Rafael Nadal (3954.51) vs. Werner Eschauer (727.6).

Table1 illustrates the correct prediction percentage obtained for the pre-match EoW margin bands across the columns and the pre-match SPARKS margin across the rows. This table forms the base of our analysis. Different aspects of the table have been analysed and results interpreted in the following sections.

3.1 Comparison Between Total SPARKS, SPARKS with negative EoW margin and SPARKS with positive EoW margin correct prediction percentages.

The EoW method was initially analysed to check if it could be incorporated and used as a measure to help improve SPARKS predictability, thus we compare the correct SPARKS prediction on percentages for all matches and compare this percentage against the correct prediction percentage for those matches where our favourite had an easier win previously.

	SPARKS with Negative Pre-match EoW margin correct prediction %					SPARKS with Positive Pre-match EoW margin correct prediction %				
Pre-match SPARKS Margin	Very tough (<-0.4)	Sig. tough (-0.31 - (-0.4))	Mod. tough (-0.21 - (-0.3))	Fairly tough (-0.11 - (-0.2))	Slightly tough (-0.01 - (-0.1))	Slightly easy (0.01 - 0.1)	Fairly easy (0.11 - 0.2)	Mod. easy (0.21 - 0.3)	Sig. easy (0.31- 0.4)	Very easy (>0.4)
>1600	–	83.33%	80.95%	80.00%	91.45%	95.21%	96.30%	90.10%	96.88%	90.24%
1281-1600	100.00%	100.00%	75.76%	83.93%	82.09%	83.54%	82.97%	85.61%	87.50%	83.33%
961-1280	66.67%	75.00%	79.69%	78.40%	77.00%	81.31%	84.10%	82.38%	82.18%	74.42%
641-960	57.14%	76.74%	66.42%	75.47%	75.39%	75.44%	70.63%	73.29%	78.45%	70.89%
321-640	68.18%	57.38%	61.80%	64.61%	64.43%	65.76%	68.17%	67.99%	71.01%	73.49%
0.1-320	46.00%	48.18%	53.49%	52.45%	57.95%	53.39%	59.07%	57.88%	59.70%	50.55%
Total win	56.18%	60.08%	61.97%	63.69%	67.88%	68.67%	72.15%	71.59%	76.03%	72.71%

Table 1: Comparison of prediction percentage of Pre-match SPARKS Margin against Pre-match EoW margin score

Looking at Table 1, 46.00% in the pre-match SPARKS margin band (0.1 – 320) plus a very tough EoW margin band can be interpreted as;
Out of all those matches where our favourite and his opponent were of similar strength/quality and where our favourite had a very tough match in the previous rounds as compared to his opponent, 46% of those matches were correctly predicted to win as per SPARKS.

Table 2 shows a comparison of correct prediction percentage where our favourites had a tougher previous match, easier previous match and for all matches in general. The quality of the player is also accounted for here.

Pre Match SPARKS Margin	SPARKS (-) pre-match EOW margin correct prediction %	SPARKS (+) pre-match EoW margin correct prediction %	Total SPARKS correct prediction %
>1600	87.71%	94.29%	92.83%
1281-1600	82.55%	84.05%	83.67%
961-1280	77.49%	81.96%	80.61%
641-960	73.91%	73.49%	73.63%
321-640	63.85%	67.63%	66.08%
0.1-320	54.81%	56.12%	55.53%
Total	65.34%	71.05%	68.92%

Table 2: Comparison of correct prediction percentages without incorporating EoW against after incorporating EoW method.

Table 2 illustrates that for closely rated or similar quality players (0.1 – 320) overall SPARKS correctly predicts the winner 55.53% of the time. The sample size for these matches was 4140 which consisted of all players who were of similar quality. Out of which SPARKS correctly predicted the winner in 2299 matches thus giving it a win percentage of 55.53%. Out of the 4140 matches there were 2288 matches where our favourite i.e. the predicted winner had an easier win in the previous round and 1852 matches where our predicted winner had a tougher win in the previous round. Thus in cases where our favourite had an easier win SPARKS predicted correctly 56.12% of the time and in cases where our favourite had a tougher win SPARKS predicted the correct winner 54.81% of the time.

Therefore indicating that every time there is a match between similar strength players there is a 55.53% chance that SPARKS will get the prediction correct. However if the predicted winner also has had an easier previous win there is a higher chance (56.12%) of obtaining the correct prediction. Thus EoW can be incorporated and accounted for when looking at SPARKS ratings to further improve its predictability.

Similarly looking up across the different SPARKS margin bands we see that in all cases the SPARKS with positive EoW outperforms the overall SPARKS result except in band (641 – 960) where it slightly underperforms in comparison. In fact on the topmost bands (>1600) where there is a severe difference in ratings between players overall SPARKS gets prediction right 92.83% of the time whereas when we considering previous performance, in particular when the favourite has an easier win this can increase the prediction by 1.46% to 94.29%. Overall prediction not accounting for the player's quality sees an increase in prediction percentage of 2.13% (71.05% - 68.92%) where the favourite has easier runs previously.

3.2 Comparison of correct prediction percentages not accounting for quality of players.

Further an in depth analysis was conducted to compare each SPARKS with positive EoW margin band against the corresponding negative bands. Table 3 shows the total matches, the number of correct predictions and the correct win percentage in each band for both the positive and negative sides. The percentage difference between each of the positive and negative bands are also calculated.

EoW Bands	SPARKS with (-) EoW (Favourite had a tougher win than opponent in previous matches)			SPARKS with (+) EoW (Favourite had an easier win than opponent in previous matches)			Percentage difference (Positive - Negative)
	Correct predictions	Total	Correct prediction %	Correct predictions	Total	Correct prediction %	
Slightly tough/ easy	1773	2612	67.88%	2168	3157	68.67%	0.79%
Fairly tough/easy	791	1242	63.69%	1891	2621	72.15%	8.46%
Moderately tough/easy	453	731	61.97%	1038	1450	71.59%	9.62%
Significantly tough/easy	149	248	60.08%	463	609	76.03%	15.95%
Very tough/easy	50	89	56.18%	341	469	72.71%	16.53%

Table 3: The total matches, the number of correct predictions and the correct win

Looking at the table we see, the cases where the favourite has an easier run in their previous match outperforms the cases where the favourite has a tougher match previously in every EoW band. In other words, there always is a higher percentage of correct predictions when the favourites have had an easier game before than those who have had a tougher game previously. Thus, we see the percentage difference in the last column between these positive and negative EoW cases is always positive. We also see that the percentage difference increases as the degree of intensity increases across the EoW bands. When the favourites have a slightly easy previous match and are compared against those that have a slightly tough match, the difference in correct prediction percentage is 0.79%.

However this increases to 8.46% when the fairly easy band is compared against fairly tough. This further increases along the rest of the bands as well. Looking at the last EoW band we see the difference between our correct prediction percentages to be 16.53%. This is because the favourites who have had a very easy match will definitely have a far higher probability of winning against those that have very tough previous matches. Thus, these results indicate that previous performance could have potential influence on the result of the subsequent match.

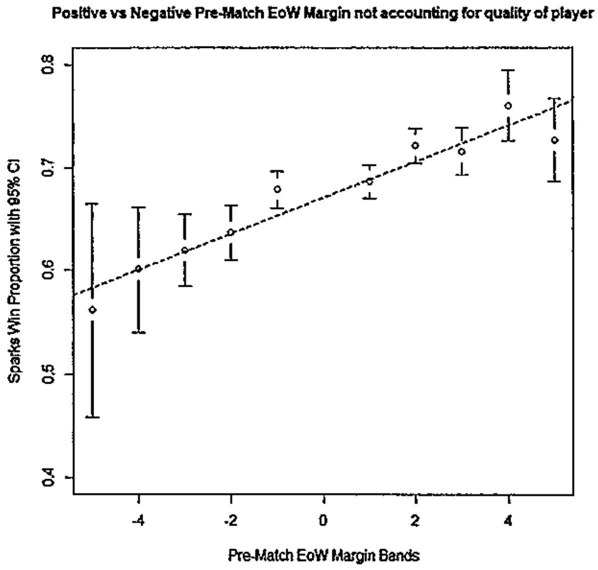


Figure 1: Positive vs. negative pre-match not accounting for quality of player

Figure 1, illustrates the win percentages in each pre match EoW band where the X axis refers to -5/-4/-3/-2/-1 as: Very/Significantly/Moderately/Fairly/Slightly tough previous win bands and 5/4/3/2/1 as Very/Significantly/Moderately/Fairly/Slightly easier previous win bands with their respective error bars taking into account the sample size in each case. A linear regression was carried out as well on the SPARKS correct prediction percentage and results indicate r^2= 0.9243. Thus verifying as the degree of toughness in previous match was reduced, the correct prediction percentage increased.

3.3 Comparison of correct prediction percentages accounting for quality of players.

Further we decided to analyse our SPARKS prediction percentage incorporating the pre-match EoW margin while accounting for the quality of player and his opponent. The smallest margin (0.1 – 320) as we know indicates similar quality/strength competitors in a match. Most of these matches are more difficult to predict. Similarly >1600 margin indicates that the two players differ greatly in strength and matches in these cases are skewed indicating a possibility of greater correct prediction percentage. Does incorporating the EoW score help increase/decrease the predictability in SPARKS irrespective of player quality or does it differ depending on different bands of player quality?

Figure 2 illustrates the prediction percentage comparing the tough/easy degrees of intensity for each band representing player quality. We see a positive trend in each case indicating that most of the time the prediction percentage for our favourite with positive pre-match EoW margin is greater than those with a negative pre-match EoW margin at different intensities.
We also see that the SPARKS correct win percentages keep increasing from the 1st to the last figure (lowest quality difference band to highest quality difference band) indicating that overall our correct prediction probability increase with an increase in the form difference between players as expected.

3.4 Comparison across each round of the tournament

Further analysis was conducted to check the Ease of Win effect across different rounds of the tournament. For this purpose, the entire data set was segregated into Round 2, Round 3, Round 4, Quarter-final, Semi-final and Final matches. Round 1 is not accounted for as there is no pre-match EoW for the players, this being the first match of the tournament. Table 4 illustrates the correct prediction percentages.

Round 2 & Round 3 have higher prediction differences (6.02% & 7.94% respectively) between the positive and negative pre-match EoW margin as compared to the later rounds. This indicates that we would be more likely to predict correctly when our favourite has an easier win as opposed to him having a difficult win in the earlier rounds of the tournament than in the latter rounds.

Round	SPARKS with (-) Pre-match EoW margin correct prediction %	SPARKS with (+) Pre-Match EoW margin correct prediction %	Total SPARKS correct prediction %	% Difference (+/-)
Finals	66.52%	72.16%	69.95%	5.64%
Semi-finals	63.02%	68.53%	66.15%	5.51%
Quarter-finals	65.03%	67.80%	66.67%	2.77%
Round 4	69.41%	74.43%	72.45%	5.02%
Round 3	64.22%	72.16%	69.43%	7.94%
Round 2	65.83%	71.85%	69.72%	6.02%
Overall	65.34%	71.05%	68.92%	5.71%

Table 4: Comparison of percentage difference between positive & negative pre-match EoW correct prediction across each round of the tournament

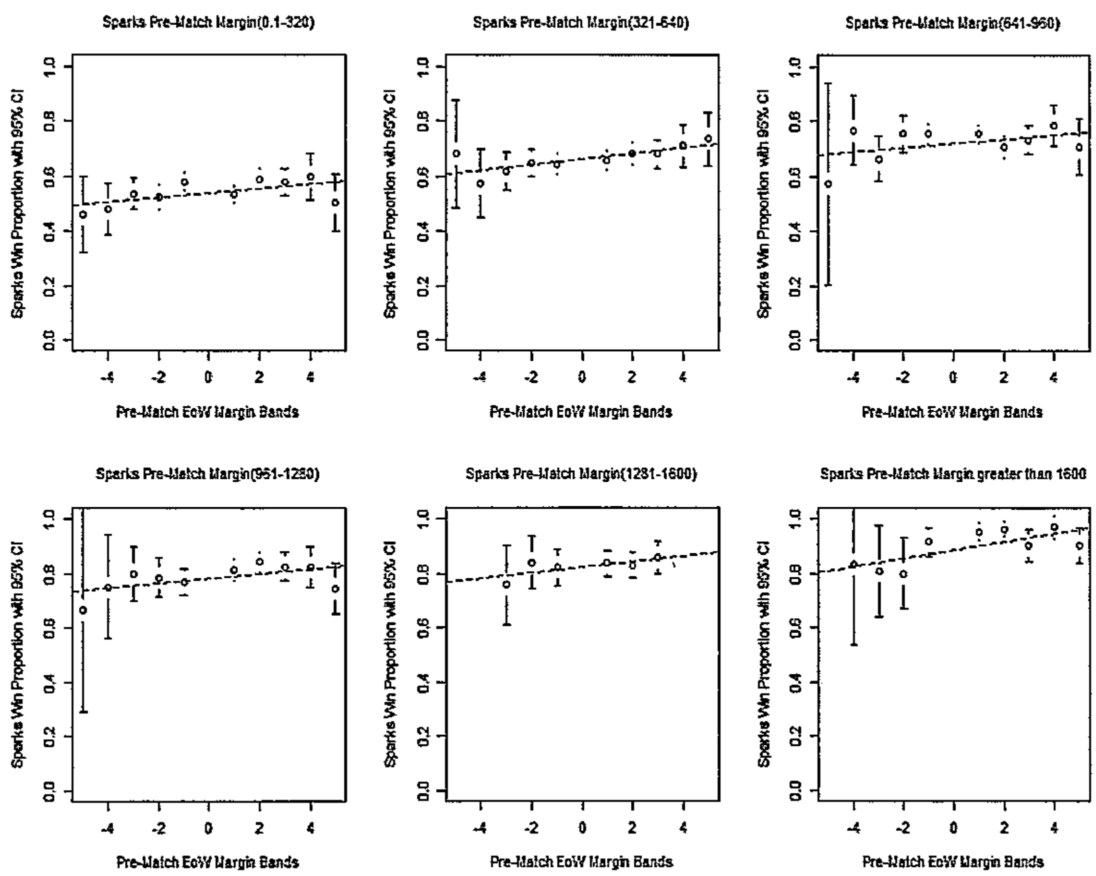


Figure 2: The prediction percentage: comparing the tough/easy degrees of intensity for each band representing player quality

4. CONCLUSION

In this study, the EoW approach is considered within-tournament using ATP data over nine years (2003 – 2011, n= 13,228). Nested EoW, which is a running average of the EoW across the rounds, was controlled for. Overall results suggest a systematic effect on the win percentages, after accounting for expected form (SPARKS residuals). When we consider EoW independently of expected form, the pre-match predicted favourites with an easier previous match win 5.71% more matches than those with a tough previous match. The results obtained strongly reinforces the hypothesis that pre-match EoW influences outcomes in the next round as there is a systematically higher likelihood of winning when there is a positive EoW. Embedding EoW into the ratings model (SPARKS) improves the models predictive results by 2.12% (71.05% vs. 68.92%). The consistent positive effect of EoW indicates it is an important variable to consider in the performance, and potential performance, of tennis players within tournaments.

REFERENCES

Von Duvillard S P, Rihacek C. et al. (2001). A physiological profile of tennis match play. Med Sci Sports Exerc , 33999–1005.1005.

Chandler T J. (1995). Exercise training for tennis. Clin Sports Med, 1433–46.46.

O'Donoghue P, Ingram B. (2001). A notational analysis of elite tennis strategy. J Sport Science, 19107–115.115.

Ferrauti A, Pluim M B, Weber K.(2001). The effect of recovery duration on running speed and stroke quality during intermittent drills in elite tennis players. J Sport Science19235–242.242.

Girard O, Millet G P.(2004). Effects of the ground surface on the physiological and technical responses in young tennis players. In: Reilly T, Hughes M, Lees A, eds. Science and racket sports III. London: E & F N Spon, 43–48.48.

Bedford,A. and Clarke, S. (2000). A comparison of the ATP ratings with a smoothing method for match prediction, in Fifth Conference on Mathematics and Computers in Sport, G. Cohen and T. Langtry (editors), University of Technology, Sydney, Australia , 43-51

APPLYING STATISTICAL TESTS FOR THE INDEPENDENCE OF POINTS IN TENNIS

Pollard, Geoff [b, c] and Pollard, Graham [a]
[a] *Faculty of Information Sciences and Engineering, University of Canberra, Australia*
[b] *Faculty of Life and Social Sciences, Swinburne University of Technology, Melbourne, Australia*
[c] *Corresponding author: gpollard@tennis.com.au*

Abstract

In the probabilistic modelling of sports such as tennis, the assumption is typically made that points are independent and thus the probability of the server winning a point is constant over a service game or set or even a match. This paper develops a range of tests to check this hypothesis and applies these tests to a top ten player in Grand Slam best of five sets matches against other top ten players.
It considers the probability of the server winning a point (a) if he is ahead, equal or behind in a game, (b) whether he won or lost the previous point, (c) given various combinations of (a) and (b), and (d) given the importance of the point. It also looks at other comparisons such as the actual and expected (assuming independence) number of games won on service, the actual and expected duration of games, the number of runs of wins and losses, and the distribution of set scores.

The analysis begins by looking at the first set of the French Open Final 2011 between Nadal and Federer, then the four set match and finally a summary of all the eleven matches and forty sets that Nadal played against other Top Ten players in the four Grand Slams in 2011. Most tests confirmed the assumption of independence, but it was found firstly that Nadal can lift when receiving and ahead in that game and secondly can lift when serving having lost the previous point, so that not all points are independent.

Keywords: Independent, score-dependent and stepwise-dependent points, importance, duration, Wald-Wolfowitz

1. INTRODUCTION

The assumption made in almost all probabilistic modelling of sports such as racket sports, volleyball, etcetera is that points are independent. This independence assumption is also made in sports like golf where holes are assumed to be independent. A range of characteristics of a scoring system such as the mean, the variance and the skewness of the number of points played, and the probability that each player wins, can be evaluated using mathematical methods. It is possible that some of these characteristics are more dependent on this assumption of independence than others. It is also possible that for some players this independence assumption is reasonable for all matches, whilst other players may exhibit signs of non-independence in some matches.
In this paper we consider several general measures that might be used in a search to see whether lack of independence might exist. Some measures are used to identify a well defined or straight forward source of possible independence, whilst other measures might identify one or more of several possible sources of non-independence. In all we consider eight measures with the view to possibly identifying which of the measures may be more effective at identifying lack of independence when it exists.

It would seem that a very good player would be able to lift his play against a not-so-good player, thus exhibiting non-independent points, so we have omitted such 'one-sided' matches from consideration. We consider one very good player and focus on his close matches where he is playing other very good players, and where he presumably would like to lift his game if he can at the more important stages in the match.

There is little published research on testing whether players do have constant probabilities of winning a point on service and that points are independent and identically distributed. Using a large panel of matches played at Wimbledon 1992-1995 Klaassen and Magnus (2001) showed that winning a point has a positive effect on winning the next point and the server finds it more difficult to win important points than less important points.

Pollard (2004) showed that Agassi was able to lift on selected points in the seven matches he played to win the Australian Open in 2003 and consequently in this case not all points are independent. Pollard, Cross and Meyer (2006) used data from Grand Slam Men's Singles matches from 1995 to 2004 to show that the pattern of ten different possible scores for five set matches could not have been achieved if the probability of winning a set was constant and consequently the better player can lift his play in certain circumstances some of the time.

2. METHODS

Many players try as hard as possible on every point, and thus are unable to lift their play on particular points. In the absence of any psychological aspects such as 'tightening up' on important points, points for such players might be expected to be independent.

Very good players might be able to lift their play on important points against average players. These are matches that the better player is typically going to win anyway.

Of particular interest is whether the very best players can lift their play against other players of the same stature. If this is the case, then the probability of winning a point is not constant. This is an aspect of lack of independence of points.

In this paper we consider one of the very best players, Nadal, and we consider matches where he may have wished to try harder at certain stages. These matches are those against other top ten players in the Grand Slam events.

The approach taken is that the type of lack of independence is a possible factor that depends on the particular player, so we focus on that player, and his matches.

We develop a range of statistical tests for identifying specific and non-specific types of lack of independence. The purpose of this paper is to delineate a set of such tests which might be used to see whether any player has such 'non-independent point characteristics'. It is not our primary purpose in this paper to determine whether this player in particular does play in a non-independent fashion. Our purpose is merely to show how such a full analysis might be carried out.

The set is the fundamental component of our analysis. The difference between winning and losing a set can be the outcome of just one or two points.

All the data comes from the IBM Pointstream analysis on the official website of each Grand Slam for 2011. Pointstream was discontinued in 2012.

The final of the French Open, 2011, Nadal vs Federer, first set

Nadal played Federer in the final of the French Open 2011 and won the first set 7-5. Federer served in the first game of the set, and won 21 of his 36 points (or 58%) on service. Nadal won 26 points out of 44 on his service (i.e. 59%). Thus, their success rates on service were essentially equal, yet Nadal won the set. In this paper we try to identify statistical measures as to why one player wins a set even though the two players are effectively equal. Using W (L) to represent a point won (lost) by the server, the outcomes on Federer's service games were

1. 0-0 W, 15-0 W, 30-0 W, 40-0 L, 40-15 W
3. 0-0 W, 15-0 W, 30-0 W, 40-0 L, 40-15 W
5. 0-0 L, 0-15 W, 15-15 W, 30-15 L, 30-30 W, 40-30 W
7. 0-0 W, 15-0 L, 15-15 L, 15-30 W, 30-30 W, 40-30 W
9. 0-0 W, 15-0 L, 15-15 W, 30-15 L, 30-30 L, 30-40 L
11. 0-0 L, 0-15 L, 0-30 W, 15-30 W, 30-30 L, 30-40 W, D L, adR L,

and the outcomes on Nadal's service games were

2. 0-0 L, 0-15 L, 0-30 W, 15-30 L, 15-40 W, 30-40 W, D L, adR W, D L, adR L
4. 0-0 W, 15-0 L, 15-15 W, 30-15 L, 30-30 W, 40-30 W

6. 0-0 W, 15-0 L, 15-15 W, 30-15 W, 40-15 L, 40-30 W
8. 0-0 W, 15-0 W, 30-0 L, 30-15 W, 40-15 L, 40-30 L, D L, adR W, D W, adS W
10. 0-0 L, 0-15 W, 15-15 L, 15-30 W, 30-30 W, 40-30 W
12. 0-0 W, 15-0 L, 15-15 L, 15-30 W, 30-30 W, 40-30 W.

It can be seen that Federer lost 2 service games whilst Nadal lost just one, with Nadal winning the set 7-5.

We now consider a range of statistical measures applied to this set.

Measure 1. State-dependent relative frequencies

We begin by noting that if a server has a probability of 0.6 of winning any point on service and the points are independent, the expected duration of the game is 6.4842 points and his probability of winning the game is 0.7357. If however his probability of winning a point is 0.7 when behind, 0.6 when equal and 0.5 when ahead in the game, the expected duration of the game increases to 7.5948 and his probability of winning the game increases to 0.7537.

In this first set, we can see that Nadal won 8 out of 16 points (50%) when ahead on his service, 10 out of 17 points (59%) when the points' score on his service was equal, and 8 out 11 points (73%) when he was behind on service. In comparison Federer won 8 out of 14 points (57%) when ahead on his service, 8 out of 14 points (57%) when the points' score on his service was equal, and 5 out 8 points (63%) when he was behind on service. Thus Nadal did a little better when behind on service than did Federer. This would appear to have been slightly to his advantage in managing to lose fewer service games than Federer in this set.

As we use the various statistical measures in this paper over several matches played by Nadal against other top ten players, we need to summarize each statistical measure.

The relevant statistical summary for this measure is: Nadal (8/16, 10/17, 8/11; 26/44) and Federer (8/14, 8/14, 5/8; 21/36).

Measure 2. Stepwise relative frequencies

It can be seen that if a player always wins the next point having lost the previous point, then he must win the set 6-0. This fact helps to motivate this second measure.

We note that if the server's probability of winning the first point in the game is 0.6, his probability of winning a point having lost the previous point is 0.7, whilst his probability of winning a point having won the previous point is 0.5, the expected duration of the game is 7.2005, and his probability of winning the game is 0.7523. Both of these values are higher than their corresponding values when the point probability is constant at 0.6. In this section we consider the two players' stepwise relative frequencies.

It can be seen that Nadal won 10 out of 21 points (48%) having won the previous point on service, and 12 out of 17 points (71%) having lost the previous point on service. In this analysis the first point in each game has been omitted from consideration for obvious reasons. Correspondingly, Federer won 9 out of 17 points (53%) having won the previous point on service, and 8 out of 13 points (62%) having lost the previous point. Thus, Nadal won a higher percentage of points having lost the previous point than did Federer. This characteristic would appear to have been to Nadal's advantage.

The relevant statistical summary for measure 2 is: Nadal (10/21, 12/17) and Federer (9/17, 8/13).

Measure 3. Combined state and stepwise relative frequencies

Suppose the server's probability of winning a point is 0.6 except when he is ahead having won the previous point when it is 0.5 and when he is behind having lost the previous point when it is 0.7. The expected duration of his service game is now 7.4026, and his probability of winning the game is 0.7521.

We now consider the stepwise relative frequencies conditional on the scores. These are given in Table 1. It can be seen that Nadal won 6 out of 13 points (46%) when ahead and having won the previous point, whereas he won 7 out of 9 points (78%) when behind and having lost the previous point. Correspondingly, Federer won 6 out of 12 points (50%) when ahead and having won the previous point, whereas he won 8 out of 13 points (62%) when behind and having lost the previous point. Again, these statistics when behind having lost the previous point (78% versus 62%) operated in Nadal's favour.

	Na	Na	Na	Fe	Fe	Fe
State	W	L	Rel F	W	L	Rel F
A,W	6	7	6/13	6	6	6/12
A,L	2	1	2/3	2	0	2/2
E,W	3	3	3/6	2	2	2/4
E,L	3	2	3/5	2	2	2/4
B,W	1	1	1/2	1	0	1/1
B,L	7	2	7/9	4	3	4/7
Tot,W	10	11	10/21	9	8	9/17
Tot,L	12	5	12/17	8	5	8/13

Table 1: Combined state and stepwise relative frequencies, Rel F for Nadal (Na) and Rel F for Federer (Fe), for the various states (Ahead (A), Equal (E), Behind (B), and previous point outcome of won (W) or lost(L))

The relevant statistical summary for measure 3 is: Nadal (6/13, 2/3; 3/6, 3/5; 1/2, 7/9) and Federer (6/12, 2/2; 2/4, 2/4; 1/1, 4/7).

Measure 4. Importance-based relative frequency method

The importance of a point within a game of tennis is defined as the probability the server wins the game given he wins that point minus the probability he wins the game given he loses that point (Morris, 1977). Table 2 gives the importances of each of the points in a game when p = 0.6, a parameter value that is relevant for many men's singles matches. The points have been listed in the order of their importances from the most important to the least. Advantage receiver, AdR includes 30-40 and advantage server, AdS includes 40-30, and 30-30 includes deuce. A close look at the 4th and 6th columns of Table 2 reveals that Nadal performed better than his average on the more important points on his service, whilst Federer performed worse than his average on the more important points on his service (note that this second observation is the same as Nadal performing better on those more important points on Federer's serve).
Correspondingly to above, it can be shown that if the server lifts his p-value to 0.7 on those points at least as important as 15-15, whilst he lowers his p-value to 0.5 on all the other points, the expected duration of the game is 7.4528 points and the server's probability of winning it is 0.7979.
The relevant statistical summary for measure 4 is: Nadal (3/4, 7/11, 9/14, 10/15, 11/16, 13/19, 15/23, 20/29, 23/35, 25/38, 25/38, 26/41, 26/42, 26/44,26/44) and Federer (1/3, 3/8, 5/10, 5/10, 6/11, 7/13, 9/16, 11/18, 15/24, 15/26, 15/26, 17/30, 19/32, 21/34, 21/36).

Score	Imp	Na	Na	Fe	Fe
		RF	Cum. RF	RF	Cum. RF
adR	0.6923	3/4	0.75	1/3	0.333
30-30	0.4615	4/7	0.636	2/5	0.375
15-30	0.4431	2/3	0.643	2/2	0.500
15-40	0.4154	1/1	0.667		0.500
0-30	0.3655	1/1	0.688	1/1	0.545
0-15	0.3456	2/3	0.684	1/2	0.538
15-15	0.3323	2/4	0.652	2/3	0.563
adS	0.3077	5/6	0.690	2/2	0.611
0-0	0.2659	3/6	0.657	4/6	0.625
30-15	0.2585	2/3	0.658	0/2	0.577
0-40	0.2492		0.658		0.577
15-0	0.2127	1/3	0.634	2/4	0.567
30-0	0.1329	0/1	0.619	2/2	0.594
40-15	0.1231	0/2	0.591	2/2	0.618
40-0	0.0492		0.591	0/2	0.583

Table 2: Importance-related relative frequencies RF and Cumulative relative frequencies for Nadal and Federer (Imp is the importance of the point)

Measure 5. Number of games won on service

The probability P that the server wins his service game when p is his probability of winning a point is given by

$$P = p^4(1-16q^4)/(p^4-q^4) \qquad (1)$$

(Kemeny and Snell, 1960, p.163) where q = 1 - p. When p = 26/44 (Nadal's relative frequency for this set), P is equal to 0.7165, and when p = 21/36 (Federer's relative frequency), P is equal to 0.6999. Thus, Nadal would have been expected to win 4.2988 games when 6 were played, and Federer would have been expected to win 4.1997 games out of six. However, Nadal actually won 5 service games out of the six whilst Federer won 4 out of 6, so it is clear that the actual arrangements of the points within his service games operated in Nadal's favour. This could have occurred by chance or maybe Nadal 'rose to the occasion' when he really needed to in one or more games (eg. on the 7th point of the 8th game).
It is noted that this approach (over many sets) might lead to a useful/powerful test for identifying varying values of p (for the server within a set). Recall that test 1 is said to be a more powerful test than test 2 if it has a higher probability (than test 2) of rejecting the null hypothesis when it is false.
The relevant statistical summary for measure 5 is: Nadal (p = 26/44, P = 0.7165, Observed = 5, Expected = 4.2988) and Federer (21/36, 0.6999, Observed = 4, Expected = 4.1997).

Measure 6. Duration of a game of tennis

a) The expected duration of a game of tennis is given by

$$\mu_1 = 4(p^4 + q^4 + 5s(p^3 + q^3) + 15s^2r^{-1} + 10s^3(3 + r)) \quad (2)$$

where s = pq and r^{-1} = 1-2pq. The second non-central moment of the duration of a game of tennis is given by

$$\mu_2 = 16(p^4 + q^4) + 100s(p^3 + q^3) + 360s^2r^{-1} + 20s^3(36 + 24r + 4r^2(1 + 2s)) \quad (3)$$

and the variance of the duration of a game of tennis is given by $\mu_2 - \mu_1^2$. These formulae agree with the numerical results in Pollard (1983).

We noted above that if the server lowers his p-value when ahead having won the previous point and lifts his p-value when behind and having lost the previous point, the expected duration of the game is increased relative to its value when the p-values are constant.

Thus, if the duration of a game is longer than expected, this could be a random occurrence and/or it could be as a result of the server lifting and lowering his point p-value. However, the change in the expected duration of a game for a moderate lifting and lowering of the p-value is quite small. Thus, given that the variance of the duration of a game is relatively large, it is clear that a test for lack of independence of points based on the actual duration will not be a powerful one.

For Nadal the estimate of p of 26/44 gives an estimated mean duration of 6.5284 and an estimated variance of 6.8698. Thus, assuming games are independent, the expected duration of the 6 games played is 39.1701 and the variance is 41.2186. Thus, the observed duration of 44 has a standardized Z-score of 0.7523.

For Federer. p = 21/36 gives an estimated mean of 6.5625 and an estimated variance of 6.9950.

The statistical summary for measure 6a is: Nadal (est E(D) = 6.5284, est Var(D) = 6.8698, number of games = 6, Observed Duration = 44, Z = 0.7523) and Federer (est E(D) = 6.5625, est Var(D) = 6.995, number of games = 6, Observed Duration = 36, Z = -0.5209).

b) We could also devise a test using just the games won by the server. This test is based on the distribution of duration conditional on winning. The expected duration conditional on the server winning his service game is

$$\mu_1 = P^{-1}(4p^4(1 + 5q + 15q^2) + 20pq^3r^2(2 - 3s))) \quad (4)$$

and the second non-central moment of the duration of a game of tennis conditional on winning is given by

$$\mu_2 = P^{-1}(16p^4 + 100p^4q + 360p^4q^2 + 20p^5q^3r^3(64 - 184s + 144s^2)) \quad (5)$$

Note that this test covers say about 75% of service games (i.e. it omits those games lost by the server).

For Nadal the estimate of p is 26/44, giving an estimated conditional mean and variance of 6.4133 and 6.7170 respectively. There were 34 points in the 5 games that Nadal won. The expected duration of the 5 winning games is thus 32.0667 and the variance is 33.5849, giving Z = 0.3336. This Z-value happens to be lower than the value above since the game that Nadal lost on service was a 'long' one.

For Federer, the values conditional on winning (when p = 21/36) are a mean of 6.4508 and variance of 48.4618.

The statistical summary for measure 6b is: Nadal (est E(D) = 6.4133, est Var(D) = 6.7170, Observed Duration = 34, number of winning games = 5, Z = 0.3336) and Federer (est E(D) = 6.4508, est Var(D) = 6.8495, number of winning games = 4, Z = -0.7266).

Measure 7. The Wald-Wolfowitz two sample runs test

Using this measure, in the case of Nadal's service games for example, we line up the 44 points in his service games, and determine the number of runs of wins and losses. The number of runs has an approximate normal distribution with mean and variance given by

$$E(R) = 1 + 2n_1n_2/(n_1 + n_2) \text{ and} \quad (6)$$

$$V(R) = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)} \quad (7)$$

where n_1 is the number of points won by the server and n_2 is the number of points lost (Siegel, 1955). This normal approximation is quite good when n_1 and n_2 are greater than 10, as is typical for a set of tennis. Thus the associated Z-test is given by

$$Z = (R - E(R)) / SQRT(V(R)). \quad (8)$$

The relevant statistical summary for measure 7 is: Nadal (n_1 = 26, n_2 = 18, E(R) = 22.2727, V(R) = 10.0292, R = 26, Z = 1.1769) and Federer (n_1 = 21 and n_2 = 15, E(R) = 18.5, V(R) = 8.25, R = 18, Z = -0.1741).

Measure 8. Distribution of set score

Given the p-values for each player and a knowledge of who served first in the set, the distribution of the set score (6-0, 6-1, 6-2, ...0-6) assuming points are independent, can be derived (see Pollard, 1983). If a player can 'rise to the occasion' on one or more particularly important points so as to win a higher than usual proportion of such points, he is likely to have a more favourable set score than otherwise.
Given Federer served in the first game of this set and his p-value is estimated at 21/36 and Nadal's p-value is estimated at 26/44, the cumulative distribution of the set score is given by

Set Score	Probability	Cum dist^n
6-0	0.0099	0.0099
6-1	0.0244	0.0343
6-2	0.0944	0.1287
6-3	0.0714	0.2002
6-4	0.1851	0.3852
7-5	0.0588	0.4441
7-6	0.0822	0.5263
6-7	0.0783	0.6046
5-7	0.0543	0.6589
4-6	0.0793	0.7382
3-6	0.1539	0.8921
2-6	0.0516	0.9437
1-6	0.0485	0.9922
0-6	0.0078	1.0000

Table 3: The cumulative distribution for the first set in the Federer/Nadal match where x-y is the resultant set score, x being Nadal's final games' score and y being Federer's final games' score.

We can see that the median score is 7-6 using these estimates of the players' p-values and the assumption of independent points and games. However, Nadal did slightly better than winning 7-6. He won 7-5, and the relevant end points (in the cumulative distribution) for this observed score of 7-5 are 0.3852 and 0.4441. This would appear to indicate that Nadal performed better than Federer 'when it really counted in the set'.
The statistical summary for measure 8 is thus (7-5, 0.3852, 0.4441).

3. RESULTS

The methods outlined in Section 2 are then applied to the remaining sets of the Nadal vs Federer Final to give a match summary. Next they are applied to the 11 matches (won 8, lost 3) and 40 sets (won 26, lost 14) that Nadal played against fellow top ten players over the four Grand Slam tournaments in 2011 and these results are now summarised.

Measure 1
Under Measure 1, Nadal's overall performance on service was (336/511, 254/413, 153/248, 743/1172) which has a Chi-Squared value of 2.14 with 2 degrees of freedom, which is not significant, indicating that there is no difference between Nadal's performance when he is serving whether he is ahead, equal or behind on points in each game. Consequently the points are independent with respect to this aspect (unless otherwise stated we use 5% as the level of significance in this paper). This suggests Nadal plays to his capacity on every point on service regardless of the score and that the points are independent.
The same analysis can be applied to matches won (275/398, 193/295, 107/156, 575/849) or matches lost (61/113, 61/118, 46/92, 168/329) or sets won (256/368, 175/265, 94/123, 525/756) or sets lost (80/143, 79/148, 59/125, 218/146). None of these showed a significant difference for Nadal winning the next point according to whether he was ahead, equal or behind in any game. This reconfirms the conclusion of the previous paragraph regardless of whether Nadal won or lost the set or match.
Alternatively, a simple sign test can be applied to the 40 sets noting whether the number of points won on service when ahead, equal or behind is above (+) or below (-) the performance for that set (any equality is split 50:50). Nadal was (19+, 21-) when ahead, (19.5+, 20.5-) when equal and (24.4+, 15.5-) when behind. Again the analysis can be applied separately to matches won or lost and to sets won and lost but none were significant. The difference in winning when behind, although not statistically significant, may be worthy of investigation with a larger sample of matches.
The above analyses are then repeated for the same matches when Nadal's various top ten opponents are serving. This equates to Nadal's performance when receiving. The collective opponents' service performance was (308/481, 246/403, 147/272, 701/1156) which has a Chi-Squared value of 7.26

with 2 degrees of freedom which is significant. Thus, when Nadal is receiving and ahead in the game the server performs worse and/or Nadal performs better on the next point than throughout the set. It would appear that Nadal can lift his game for the next point when receiving and ahead.

Again, the analysis can be performed for matches won by the server (ie Nadal lost) (81/113, 61/103, 43/75, 185/291) or matches lost (Nadal won) (227/368, 186/300, 104/197, 518/865) or sets won (Nadal lost) (137/200, 96/133, 34/60, 267/391) or sets lost (Nadal won) (171/281, 150/272, 113/212, 434/765). The respective Chi-Squared values with 2 df are 5.35, 4.98, 5.27 and 3.29 but none are significant.

The sign test can also be applied to the 40 sets as before and the servers were (21.5+, 18.5-) when ahead, (22+, 18-) when equal and (18+, 22-) when behind. None of the sign tests with Nadal receiving were significant, but again the analysis may benefit with a larger sample of matches.

Measure 2

Under Measure 2, which excludes the first point of each game, Nadal's performance on service is (268/491, 253/388) giving Chi-Squared of 10.2 with 1 degree of freedom which is highly significant. Nadal wins significantly more points on service after losing the previous point ($p = 0.6521$) than after winning the previous point ($p = 0.5459$).

This difference is maintained in matches won by Nadal (200/349, 179/257) with a Chi-Squared with 1df of 9.7 and in sets won by Nadal (174/310, 160/220) with a Chi-Squared with 1 df of 15.3 but not in matches (Chi-Squared 1 df of 2.05) or sets (Chi-Squared 1 df of 1.26) lost by Nadal.

The same analysis with his opponents serving and Nadal receiving (335/561, 239/404) is not significant and nor are the subsets for matches won and lost or sets won and lost.

A simple sign test over the 40 sets confirms the above as Nadal won more points after losing a point in 25.5 sets compared to winning more points after winning a point in 14.5 sets. When receiving, the equivalent figures for Nadal were 20.5 and 19.5.

Measure 3

Having found firstly that Nadal can lift when receiving and ahead in that game and secondly can lift when serving and lost the previous point (but not under other score or point situations) we now consider the joint probabilities for score (up, equal or down) and previous point (win or lose) under the third measure.

Under measure 3 Nadal's performance while serving is (284/421, 51/71, 71/116, 64/105, 18/40, 139/214) giving a Chi-Squared with 5 df of 11.11 which is just significant, while his opponents' serving performance and his receiving performance is (257/411, 46/67, 52/100, 70/110, 28/46, 121/220) giving Chi-Squared with 5 df of 8.8 which is not significant.

Although just significant at 5%level (11.11 compared to 11.07) most of the contribution to the calculated value comes from the category with the smallest frequency, namely serving when behind having won the previous point, where his observed was well below expectation. Further analysis is required before coming to a conclusion that there is any significance in these results.

Measure 4

For simplicity and data quantity considerations when assessing Nadal's performance on important points we summarise and use the same categories as in Section 2, viz, scores of 0-15 or more important as in Table 2 and scores of 15-15 or less important.

Nadal won 223 out of 375 ($p=0.595$) more important points and 493 out of 763 ($p=0.646$) less important points, but the difference is not significant (Chi-Squared 1 df is 2.7). Likewise, there was no significant difference when his opponent was serving and Nadal receiving.

Individual comparisons between specific interesting scores (e.g. love-30 versus 30-love and 40-30 versus 30-40) showed no significant difference in performance, although the comparison for Nadal at 40-30 ($p=48/62=0.77$) with 30-40 ($p=21/36=0.58$) is right on the 5% level with Chi-Squared 1 df equal to 3.89.

Measure 5

Over the 40 sets played Nadal won more games than expected (assuming his probability of winning a point on service is constant throughout that set) on 28 occasions and less games on 12 occasions which is significant. On the other hand when Nadal was receiving, his opponent won more games than expected on 18 occasions, less games on 19 occasions, and equal to expectation (assuming constant probability throughout the set) on 3 occasions. This is clearly not significant ($p > 0.05$).

This suggests that Nadal can lift at times on service so his probability of winning a point on service is not constant throughout the set. However, when receiving there is no evidence under this test of any variation during a set in the probability of his opponent winning a point on service.

Measure 6
In 3 of the 40 sets included in this analysis using Measure 6(a), Nadal played a significantly ($z > 1.96$) greater number of points on service than expected while in the other 37 the difference was not significant ($p < 0.05$). In 18 sets the difference between the observed and expected number of points was positive and in 22 sets it was negative, and this is not significant ($p > 0.05$). However, when Nadal was receiving, his opponent had more points than expected in 12 games and less than expected in 28 games, and this is significant ($p < 0.05$).
Similar results are obtained if you just consider (Measure 6(b)) the number of points played in service games won by Nadal. When Nadal served the number of points played exceeded the expected on 17 occasions and was less on 23 occasions, which is not significant ($p > 0.05$). When Nadal received, his opponent (in service games won) had less points on service than expected on 10 occasions and more than expected on 30 occasions and this is significant ($p < 0.05$).

Measure 7
In one set Nadal had a significantly greater number of runs than expected ($z = 2.15$), in another set it was less than expected ($z = -2.13$) and the other 38 sets were not significant ($p < 0.05$). In 23 sets there were more runs than expected and in 17 sets there were less runs than expected, and this is not significant ($p > 0.05$).
Similar non significant results were obtained when Nadal was receiving. The Wald-Wolfowitz two sample runs test thus did not detect any significant difference in the number of runs expected if all points in a set were independent and the probability of winning a point on service remained constant throughout the set.

Measure 8
Over the 40 sets played Nadal's average set score performance had probability limits (0.3506, 0.5722) and an average of 0.4614. This is less than 0.5 suggesting Nadal performed slightly better than expected, but the difference is not significant. His set score was better than expected (under the assumption of independence and constant probabilities of the server winning points on service) in 25 sets and worse in 15 sets ($p > 0.05$). This was also confirmed with a t-test producing a value of 1.47 with 39 degrees of freedom.

4. DISCUSSION

It is known that the percentage of points won on service by each player in a set are biased estimators of the probability that each player wins a point on service (Pollard et al., 2010). For example, they showed that the expected value of the proportion of points won on service by a server with a p-value of 0.6 was 0.6264. For a set, the winner's proportion of points won on service has a positive bias and the loser's proportion of points won on service has a negative bias. For example, when pa = pb = 0.65 and player A served in the first game of a tiebreak set, the (set) estimate of pa, pa^ averaged 0.6535, and pb^ averaged 0.6501, the proportion of points won by the winner averaged 0.7053 and the proportion of points won by the loser averaged 0.5984 over 4 million simulated sets. Also, they showed that when pa = 0.7 and pb = 0.6 and player A served first in the set, pa^ averaged 0.7103, pb^ averaged 0.5900, the proportion of points won by the winner averaged 0.7225 and the proportion of points won by the loser averaged 0.5799 over 4 million simulated sets.
Suppose T is an estimator of the parameter θ. Then, the bias in the estimator T is given by

$$Bias(T) = E(T) - \theta . \quad (9)$$

The mean square error of T is defined by

$$MSE(T) = E((T-\theta)^2) . \quad (10)$$

It can be shown that

$$MSE(T) = Var(T) + Bias(T)^2 . \quad (11)$$

Since each player must have at least 3 service games in a set, and each service game must have at least 4 points, the first 4 points in the first 3 service games for each player constitute 12 points that provide an unbiased estimate for that player's p-value.
It appears that it is not possible to produce an unbiased estimator based on more than just these 12 points. Other positively and negatively biased estimators can be produced, and one could produce (correlated) combinations of these with very little bias, but this has not been pursued for reasons that become clear below. [An example of a negatively biased estimator of p in a game is the ratio of the number of points won by the server divided by the number of points played in the game over the

restricted set of points (0-0, 15-0, 0-15, 15-15, 30-15). For example, when p = 0.6, the expected value of the proportion of points won by the server over this set of points (not all of which can occur) can be shown to equal 0.592.]

An estimate based on the first 4 points in the first 3 games can in practice provide an *inflated* estimate of p. Sometimes when the server is up 40-0 (and even 30-0) the receiver prefers to 'save his energy' for later in the match, and adopts a 'more risky' strategy, lowering his probability of winning such points. The effect of this is to inflate the server's estimated p-value (even though from a purely statistical perspective the estimate is unbiased).

The variance of the proportion of points won out of 12 points when p = 0.6 is equal to 0.02, whereas the variance of the proportion of points won out of 24 points when p = 0.6 is equal to 0.01. (Note that in almost all sets each player serves more than 24 points). If estimator 1 of a p-value is based on 24 points and has a bias of 0.05, then its MSE is equal to 0.01 + 0.0025 = 0.0125 (when p = 0.6). If estimator 2 of a p-value is based on 12 points and has a bias of zero, then its MSE is equal to 0.02. In such a situation the biased estimator 1 would be preferable. Thus, in our situation, it would appear to be preferable to use an estimator based on the larger sample size (the whole set), even though it is a biased estimator. It is clear that, in our case, the square of the bias in the above equation is somewhat dominated by the variance of the estimator.

Given that the p-estimate for the winner of the set (based on all of his service points in the set) has a positive bias (i.e. on average it over-estimates the winner's p-value), and the p-estimate for the loser of the set (based on all of his service points in the set) has a negative bias (i.e. on average it under-estimates the loser's p-value), we would expect that the median set score in measure 8 would indicate that the winner of the set should (averaged across many sets) win 'easier' than actually observed.

5. CONCLUSIONS

Given that the mathematical analysis of tennis is generally based on the assumption that points in tennis are independent and identically distributed, an analysis of this assumption is always interesting. This paper suggests eight ways to investigate the assumption and then looks at applying the analysis to real data between players of approximately equal standard.

The analysis and interpretation from a tennis and statistical point of view is first carried out for the first set of the 2011 French Open final between Nadal and Federer, then for this four set match, and finally for the eleven matches and 40 sets that Nadal played against other Top Ten players in the four Grand Slams in 2011.

While most of the results for each of the measures were not significant at the 5% level, it was found firstly that Nadal can lift when receiving and ahead in that game and secondly that Nadal can lift when serving after losing the previous point.

Consequently, even when matches are played between the best players in the world over five sets in Grand Slams and each could be assumed to be playing to his capacity on every point, some evidence that not all points are independent can be found. However, the differences, although significant, are small and the assumption of independence is a reasonable approximation.

References

Kemeny, J. G. and Snell, J. L. (1960). Finite Markov Chains, Princeton, New Jersey, D. Van Nostrand.

Klaassen, F. J. G. M. & Magnus, J. R. (2001). Are points in tennis independent and identically distributed? Evidence from a dynamic binary panel data model. *Journal of the American Statistical Association*, 96, 500-509.

Morris, C. (1977). The most important points in tennis. In *Optimal Strategies in Sports*, edited by S. P. Ladany and R. E. Machol, 131-140. Amsterdam:North-Holland. (Vol 5 in Studies in Management Science and Systems).

Pollard, G. H. (1983). An analysis of classical and tie-breaker tennis. *Australian Journal Statistics, 25(3)*, 496-505.

Pollard, G. H. (2004). Can a tennis player increase the probability of winning a point when it is important? In *Proceedings of the Seventh Australasian Conference on Mathematics and Computers in Sport*, edited by R. H. Morton & S. Ganesalingam, Massey University, Massey, New Zealand, 253-256.

Pollard, G. H., Cross, R. & Meyer, D. (2006). An analysis of ten years of the four grand slam men's singles data for lack of independence of set outcomes, *Journal of Sports Science and Medicine*, 5, 561-566.

Pollard G. H., Pollard G. N., Lyle I. and Cross R. (2010). Bias in Sporting Match Statistics. *Proceedings of the Tenth* Australasian *Conference on Mathematics and Computers in Sport*, Darwin, July, 2010, edited by A. Bedford and M. Ovens, 221-228.

Siegel, S. (1956). Non-parametric Statistics. McGraw Hill.

FAME AND FORTUNE IN ELITE TENNIS

[a,c]Denny Meyer and [a,b]Geoff Pollard

[a]*Swinburne University of Technology*

[b]*Tennis Australia*

[d]*Corresponding author: dmeyer@swin.edu.au*

Abstract

Previous research has suggested that prize money serves as an incentive for player performance, whereas rankings are considered more useful for administrative purposes such as tournament selection, seeding, handicapping and for predicting match outcomes. However, for top players it may be that it is their ranking which serves as their incentive to perform, because extraneous income and reputation is tied to rankings rather than to prize money per se. There is of course a relationship between rankings and prize money, but the intricacies of the ATP ranking system means that this relationship is perhaps not as predictable as one might expect. This paper investigates the relationship between rankings and prize money and performance using data for the top 108 ATP singles players in 2011. Performance is measured in terms of the proportion of matches won and the number of matches played in the years 2004 to 2010. Generalized linear models and more accurate multi-level models, with binomial and Poisson response distributions, are used for this purpose. Differing results for these two models indicate that it is essential to take account of the repeated measures nature of the data as is done in the multi-level analysis. However, marked differences in the results obtained for the two performance measures suggest that more work is needed in order to reach a deeper understanding of the factors motivating elite tennis players.

Keywords: rankings, prize money, multi-level models, performance incentives

1. INTRODUCTION

The use of prize money as an incentive is of course not confined to tennis. As explained by Predergast (1999), workers commonly compete for promotion in order to win the prize of a higher salary. However, in the work situation the "prize money" is not the whole story, because with a promotion comes other advantages such as increased control, power and respect. In this paper we try to establish whether prize money is really a meaningful incentive for tennis players once they reach the highest levels of their profession, or if it is "fame" factors that have more importance.

Sunde (2009) has investigated whether higher prize money improves performance in terms of the proportion of matches won and the total number of matches played. Using the Association of Tennis Players (ATP) singles data from 156 Grand Slam and Masters tournaments, collected between 1990 and 2002, Sunde showed that the higher stage prizes, and in particular the substantially higher prizes won in the finals as compared to the semifinals, significantly improved performance.

An earlier study by Gilsford and Sukhatme (2003), using Women Tennis Association (WTA) data collected from 1997 to 2004, tried to establish what factors affect the probability that the stronger player wins the match. In particular they found that the prize differential for any match, measured as the difference between the tournament's top prize and the loser's prize money associated with the current match, was a good predictor of a match win for the favoured player. Reports from the players confirm this view. For example when Wimbledon started matching the salaries of men and women players in 2007, Venus Williams was reported as saying, "The 2007 Championship will have even greater meaning and significance to me and my fellow players". But of course the prize money is not the only advantage of doing well in a tournament. Tennis rankings also provide an important incentive for players as explained below.

Ranking systems are a relatively new development in tennis. In 1973 the leading male tennis players formed their own union, the Association of Tennis Professionals (ATP), and one of their first acts was to introduce a 12 month weighted moving average system for computing rankings. This system was used to determine fairly which players should gain entry into tournaments worldwide and to determine which players should be seeded. Under the original ranking systems tournament importance or "strength" was determined by prize money and player performance was measured by the round reached.

The current ATP ranking system basically calculates how many points a player has earned in the last 12 months, particularly in 18 ATP tournaments, with rankings assigned accordingly. Points earned differ between tournaments and depend on how far a player progresses. The four Grand Slam tournaments and the eight Masters 1000 tournaments contribute points as well as a player's best four ATP 500 and best two ATP 250 events. The following ranking points are assigned to the singles champion at these 18 tournaments:-

- 2000 for a Grand Slam
- 1000 for a Masters 1000
- 500 for an ATP 500
- 250 for an ATP250

This means that players need to consistently do well in the most prestigious tournaments in order to achieve a good (low) ranking.

Rankings are very important to players for all sorts of reasons. In particular the players with the best rankings are often invited to play in special events or exhibition matches, giving them the opportunity to increase their income substantially. Very lucrative advertising contracts are another source of income for these top players. In addition a special formula is used to construct the rankings for the top eight male players. The top eight players who qualify for the year-end ATP Tour Finals (and any player who plays as an alternate) receives a bonus equal to the points they earn in the ATP Tour Final, with 1500 ranking points for the champion of this tournament.

Rankings of course have other important functions for tournament selection, seeding and handicapping and for predicting match outcomes. Using Wimbledon singles data 1992-95, Klaassen and Magnus (2003) converted player rankings into the probability that a particular player would win a match between two players. They then recalculated that probability as the match progressed. For further probability predictions based on rankings see Bedford and Clarke (2000) and Clarke and Dyte (2000).

It seems therefore that both prize money and rankings contribute to the performance of elite tennis players. In this paper we try to determine the relative importance of prize money and rankings in this regard.

2. METHODS

In this paper statistical analyses have been performed using the data provided in the ATP World Tour Media Guide (2011) for 108 singles players who competed in the 2011 ATP World Tour. The data includes the number of matches won and lost for each player in each year of their career until 2010. This data has been supplemented with On Court (2012) data for rankings and prize money for each player in each year between 2004 and 2010. The rankings and prize money data were log transformed in order to reduce the effect of outliers and to create something closer to a normal distribution as shown in Figures 1 and 2 below.

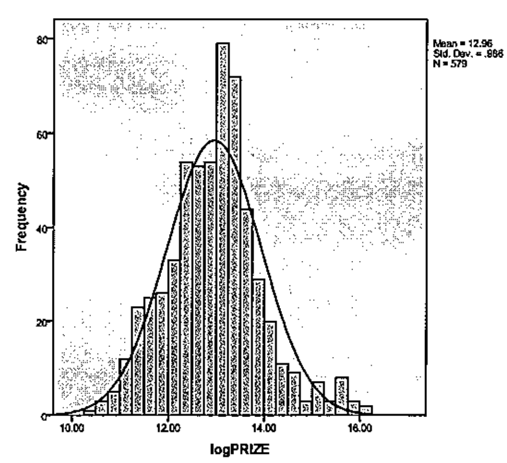


Figure 1: Log Transformed Prize Money Per Annum

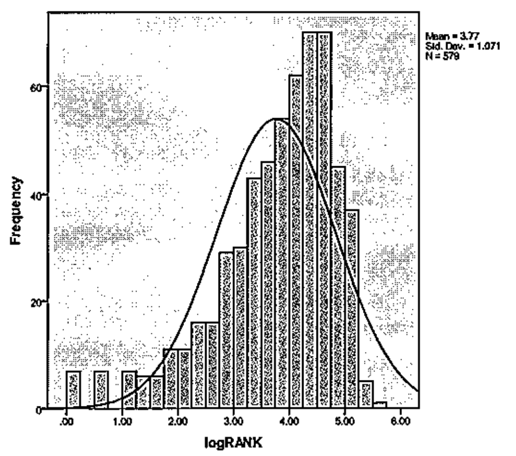


Figure 2: Log Transformed Rank

A plot of the log transformed prize and rank data in Figure 3 suggested a strong linear relationship between these variables (with r = -0.92). However, there appears to be an increase in the variation of prize money at higher rankings. This high variation in prize money at worse (higher) rankings suggests either that performance is more erratic for these players or that there is more variation in the choice of tournaments, with some players choosing tournaments with larger prizes and some players choosing tournaments with smaller prizes.

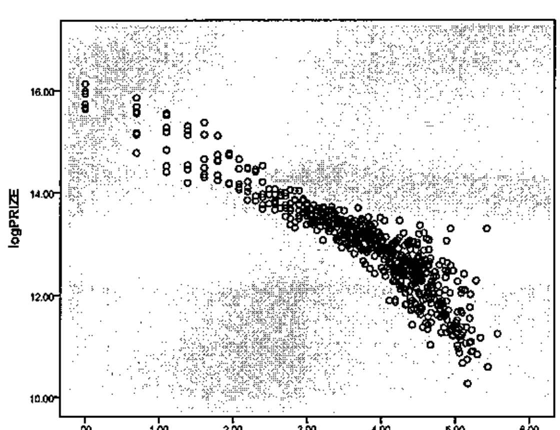


Figure 3: Relationship between Annual Prize Money and Rank after Log Transformation

Interestingly the ITF has recently introduced a compulsory International Player Identification Number (IPIN) and an associated online player entry system that in 2009 covered tens of thousands of players in well over one thousand tournaments on the ITF Junior Circuit, ITF Pro Circuit and ITF Seniors Circuit. This system allows players to easily select their most appropriate tournament at any time according to their ranking and the entry decisions of other players. If players take the advice of this system it is likely that the relationship between prize money and rankings will strengthen in the future.

As recommended by Sunde (2009). in the initial analysis conducted in this paper the proportion of matches won in any year was used as the measure of performance while in the second analysis the number of matches played in any year was used as the measure of performance. A binomial distribution with a logistic link function was assumed for the proportion of matches won in any year and a Poisson distribution with a log link function was assumed for the number of matches won in any year.

The analysis considered an average of 5.2 years of data for each of the 108 players, resulting in 566 player years of data. The average number of matches per player per year was 41 with a standard deviation of 22 while the average percentage of matches won per player per year was 51.4% with a standard deviation of 15.3%. Rankings varied between 1 and 262 with a median value of 54, and prize money won in any year varied from 29 thousand dollars (US) to 10 million dollars (US) with a median of $445000 approximately. It should be noted that during the period 2004-2010 there was a steady increase in prize money, especially for the big tournaments and for the tournament winners. This is no longer the case with players knocked out in earlier rounds now receiving a greater share of the prize money.

Initially generalized linear models were fitted, effectively assuming independence between player performance across the years. A more accurate multi-level analysis was then conducted with the results confirming that it was crucial to allow for the repeated measures nature of the data. The multi-level models are presented in Figure 4 and 5. These models allow different β coefficients for each player. They assume that on average the β coefficients for the level 1 (annual) data stay the same for each player over his career. This means that the incentive effects of rankings and prize money are assumed to remain the same over time for each player.

In Figure 4 $\square_{ij}$ represents the expected proportion of matches won by player *j* in year *i* and in Figure 5 λ_{ij} represents the expected number of matches played by player *j* in year *i*. In the level 1 models the LOGRANK and LOGPRIZE variables are group centred. This means that the β coefficients are measuring the incentive effects of LOGRANK and LOGPRIZE relative to each players mean values for these variable during the period 2004 to 2010.

Level-1 Model

$E(WIN_{ij}=1|\beta_j) = \square_{ij}{}^{*}\ MATCHES_{ij}$
$\log[\square_{ij}/(1 - \square_{ij})] = \eta_{ij}$
$\eta_{ij} = \beta_{0j} + \beta_{1j}{}^{*}(LOGRANK_{ij}) + \beta_{2j}{}^{*}(LOGPRIZE_{ij})$

Level-2 Model

$\beta_{0j} = \gamma_{00} + u_{0j}$
$\beta_{1j} = \gamma_{10} + u_{1j}$
$\beta_{2j} = \gamma_{20} + u_{2j}$

Level-1 variance = $1/[MATCHES^{*}\square_{ij};(1-\square_{ij})]$

Figure 4: Multi-Level Model for Expected Proportion (ϕ) of Matches Won

Level-1 Model

$E(MATCHES_{ij}|\beta_j) = \lambda_{ij}$
$\log[\lambda_{ij}] = \eta_{ij}$
$\eta_{ij} = \beta_{0j} + \beta_{1j}{}^{*}(LOGRANK_{ij}) + \beta_{2j}{}^{*}(LOGPRIZE_{ij})$

Level-2 Model

$\beta_{0j} = \gamma_{00} + u_{0j}$
$\beta_{1j} = \gamma_{10} + u_{1j}$
$\beta_{2j} = \gamma_{20} + u_{2j}$

Level-1 variance = $1/\lambda_{ij}$

Figure 5: Multi-Level Model for Expected Number of Matches Played (λ)

Mahalanobis Distance tests showed a Chi-Square Distribution for both the multi-level models, confirming that there are no outliers in the data. Finally Exploratory Analyses were used to check whether player characteristics such as weight, height, handedness, type of backhand and the age at which players turn professional has a significant influence on the β coefficients. None of these characteristics were found to be significant so the above simple models could be used for the multi-level analyses.

3. RESULTS

As shown in Table 1 the initial generalized linear model analysis found that the proportion of matches won in any year had a significant relationship with ranking, with lower rankings associated with better performance. However, the effect of prize money was not significant after rankings had been taken into consideration.

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	3.086	.7660	1.584	4.587	16.228	1	.000
logRANK	-.548	.0383	-.623	-.473	204.531	1	.000
logPRIZE	-.073	.0481	-.167	.021	2.292	1	.130
(Scale)	a						

Table 1: Generalized Logistic Regression for Proportion of Matches Won in Any Year

A second analysis was then performed taking into account the nested nature of the data in a multi-level (repeated measures) analysis. The results of this more accurate analysis are shown in Table 2, suggesting a significant negative relationship between prize money and performance once rankings are taken into account. This seems to suggest that as prize money increases it serves to reduce performance instead of increasing performance as was expected. Alternatively it may mean that some players deliberately choose tournaments with lower prize money so that they can increase their chances of winning or for other reasons, such as preferred court surface, location, tournament reputation.

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. $d.f.$	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	0.027982	0.045647	0.613	107	0.541
For LOGRANK slope, β_1					
INTRCPT2, γ_{10}	-0.497375	0.044098	-11.279	107	<0.001
For LOGPRIZE slope, β_2					
INTRCPT2, γ_{20}	-0.094070	0.045139	-2.084	107	0.040

Fixed Effect	Coefficient	Odds Ratio	Confidence Interval
For INTRCPT1, β_0			
INTRCPT2, γ_{00}	0.027982	1.028377	(0.939,1.126)
For LOGRANK slope, β_1			
INTRCPT2, γ_{10}	-0.497375	0.608125	(0.557,0.664)
For LOGPRIZE slope, β_2			
INTRCPT2, γ_{20}	-0.094070	0.910219	(0.832,0.995)

Table 2: Multi-Level Logistic Regression for Proportion Matches Won in Any Year

As shown by the Mahalanobis Distances in Figure 6, a Chi-Square distribution describes the distances for each player from the centroid well, suggesting that there are no outliers in the data.

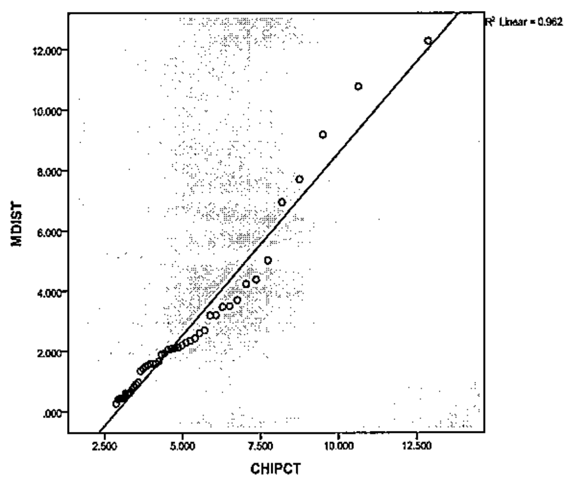


Figure 6: Check for Outliers for the Multi-Level Model for Proportion of Matches Won

Figure 7 provides darker shading when players win a higher percentage of matches in any year. The white area in the bottom lefthand corner is empty because best players always win reasonable prize money. This figure confirms the multi-level model results with the effects of ranking obviously much more important than the effect of prize money on the percentage of matches won by a player in any year. Players with the best (lowest) ratings have the highest percentage of matches won. It is really only in the case of higher (worse) ranking players that prize money seems to have much relationship with the percentage of matches won.

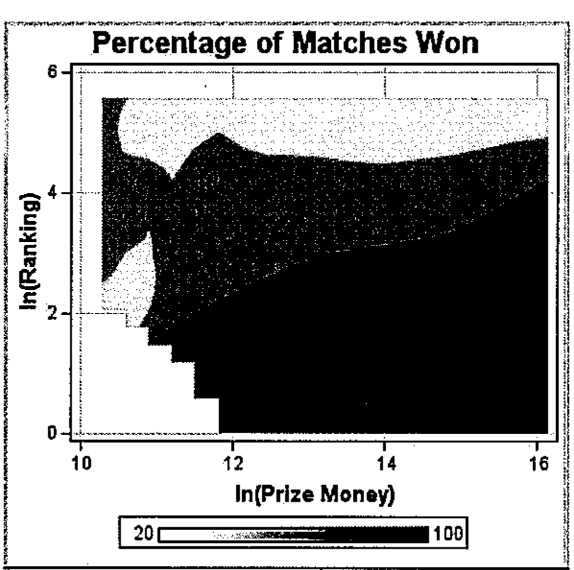


Figure 7: Contour Map for Percentage of Matches Won by a Player in Any Year.

Figure 7 suggests a cluster of players with worse (higher) rankings and low prize money who have a relatively high percentage of matches won, perhaps accounting for the negative relationship between prize money and proportion of matches won which was suggested in Table 2.

The same analyses were then performed using the number of matches played in any year as the measure of performance, assuming a Poisson distribution for this variable with a log link function. As shown in Table 3 the expected number of matches decreases for the best ranked players when prize money is controlled, but the number of matches increases when prize money increases when the ranking is controlled.

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval Lower	Upper	Hypothesis Test Wald Chi-Square	df	Sig.
(Intercept)	-5.649	.3094	-6.255	-5.043	333.434	1	.000
logRANK	.178	.0159	.146	.209	124.312	1	.000
logPRIZE	.662	.0193	.624	.700	1176.648	1	.000
(Scale)	a						

Table 3: Poisson Regression for Number of Matches Played in Any Year

The Multi-level analysis shown in Table 4 confirms this result, however for this more accurate analysis, the coefficient for LOGRANK is barely significant suggesting that the prize money earned by a player is a much better predictor of the number of matches played than rankings.

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. $d.f.$	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	3.334481	0.061901	53.868	107	<0.001
For LOGRANK slope, β_1					
INTRCPT2, γ_{10}	0.080042	0.039414	2.031	107	0.045
For LOGPRIZE slope, β_2					
INTRCPT2, γ_{20}	0.847887	0.064618	13.122	107	<0.001

Fixed Effect	Coefficient	Event Rate Ratio	Confidence Interval
For INTRCPT1, β_0			
INTRCPT2, γ_{00}	3.334481	28.063802	(24.822,31.729)
For LOGRANK slope, β_1			
INTRCPT2, γ_{10}	0.080042	1.083332	(1.002,1.171)
For LOGPRIZE slope, β_2			
INTRCPT2, γ_{20}	0.847887	2.334709	(2.054,2.654)

Table 4: Multi-Level Poisson Regression for Number of Matches Played in Any Year

The Figure 8 plot of Mahalanobis Distances for this model confirms that there are no outliers in this analysis either.

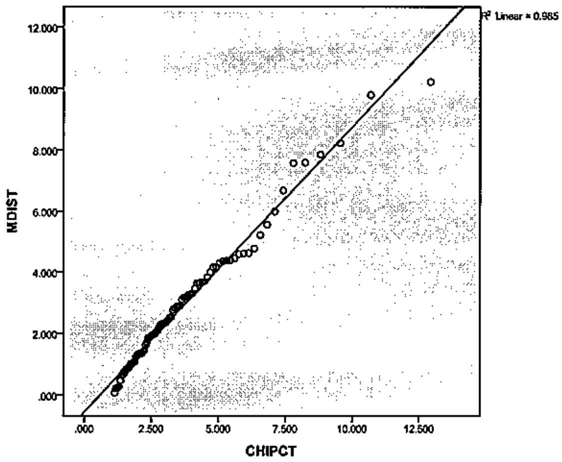


Figure 8: Check for Outliers for the Multi-Level Model for Number of Matches Played

Figure 9 shows darker shading when players play more matches in any year. As before the white area in the bottom lefthand corner shows that none of the best players received very low prize money in any year. Figure 9 confirms the results of the multi-level analysis with the effect of prize money on the number of matches played per player in any year stronger than the effect of ranking. Clearly the players with the best rankings and the highest prize money are working very hard to maintain their position on the ladder.

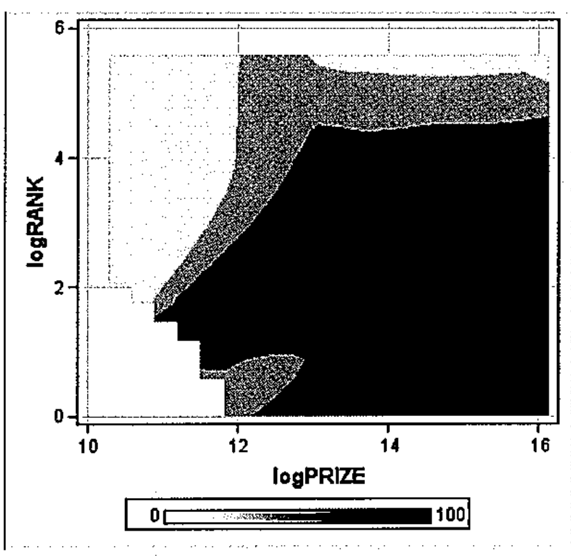


Figure 9: Contour Map for Number of Matches Played by any Player in a Year

But when prize money is low rankings appear to have little impact on the number of matches played.

One possible explanation for this relates to seeding. A player may choose to enter a smaller tournament with lower prize money in order to ensure seeding. This affords some protection in the first round since seeded players do not meet in the first round of these tournaments.

4. DISCUSSION

From the above results it appears that the importance of prize money as a predictor of performance depends on which measure of performance is used. In the case of proportion of matches won in any year, rankings are the better predictor of success, but, for the number of matches played in any year, prize money is a better predictor of success.

The negative relationship between the proportion of matches won and prize winnings when rankings are controlled is of particular interest. This suggests that there is a cluster of players with higher (poorer) rankings who are choosing to play in tournaments where prize money is relatively low, perhaps so that they will be able to increase their chances of winning because they will be meeting opponents with similar (or lower) rankings, allowing them to climb the ladder faster. For example, in the week before the Australian Open there are tournaments held in both Sydney and Hobart for women and in both Sydney and Auckland for men, with Sydney the stronger tournament with more prize money on offer. Winning a few matches at the weaker tournament will result in more rating points than a first round loss at the Sydney tournament. Also for weaker players it provides more practise before the Australian Open than the Sydney tournament, explaining why weaker players sometimes choose the tournament with lower prize money.

In addition the relatively low number of matches played by players with little prize money suggests that players in this category may need additional financial incentives if they are to progress.

However, there are some serious limitations attached to this study. Firstly the effect of increasing prize money over the period 2004-2010, with a greater share going to tournament winners later in this period, may have caused some bias.

Secondly there is likely to be some bias in the players selected for study because only current players included in the 2011 ATP World Tour Media Guide were considered. This means that the results relate only to elite men's single players. A future study will consider elite women players where different results are expected because the WTA ranking system is less stringent, favouring quantity rather than quality to a greater extent, with Grand

Slams contributing a greater percentage of the prize money. In addition the women are much more likely to play doubles than the men and it is not known what impact this will have on the results.

Finally, some thought needs to be given to using winning percentage and number of matches as response variables. These were intended as quality and quantity measures of performance, but a combination of these variables might have been a more useful way to measure performance for this study.

5. CONCLUSION

The above analysis suggests that there is no simple answer to the question concerning the relative importance of prize money and rankings as incentives for performance in elite tennis. The results of the multi-level modelling indicate that to some extent the answer to this question differs for each player. Furthermore the very different results obtained when predicting the number of matches played per year and the percentage of wins per year suggests that this question is more complex than originally appeared, requiring further investigation.

Acknowledgements

We wish to thank the International Tennis Federation and KAN-soft for the use of their data. Also thank you to the referees for their insightful comments.

References

Bedford, A.B., & Clarke, S.R. (2000), A comparison of the ATP rating with a smoothing method for match prediction. Pp 43-51 in Cohen, G, Langtry, T eds. *Proc. 5th Australasian Conference Maths and Computers in Sport.*

Clarke, S.R., & Dyte, D. (2000), Using official ratings to simulate major tennis tournaments, *International Transactions in Operational Research 7*, 585–594.

Gilsford, K.F. & Sukhatme, V.A. (2008). Tournament incentives and match outcomes in women's professional tennis. *Applied Economics, 40*, 2405-2412.

International Tennis Federation (2011). ATP World Tour Media Guide. www.ATPWorldtour.com.

Klaassen, F.J.G.M., & Magnus, J.R. (2003), Forecasting the winner of a tennis match. *European Journal of Operations Research; 148*: 257-267

On-Court (2012), www.oncourt.info, Copywright © 2001-2012 KAN-soft.

Prendergast, C. (1999). The provision of Incentives in Firms. *Journal of Economic Literature*, 37(1), 7-63.

Sunde, U. (2009). Heterogeneity and performance in tournaments: a test of incentive effects using professional tennis data. *Applied Economics, 41*, 3199-3208.

USING RELATIVE PHASE OF POSITION DERIVATIVES TO QUANTIFY WITHIN-TEAM PLAYER COUPLINGS

Stuart Morgan [a,c] and Morgan Williams [b]

[a] *Australian Institute of Sport, Australia*
[b] *University of Glamorgan, Wales*
[c] *Corresponding author: stuart.morgan@ausport.gov.au*

Abstract

Relative phase has been used to describe movement coordination between agents in game play across several sports. Notably, McGarry et al (1999) introduced the concept of phase relations to intra-player couplings in squash. Lames (2006), demonstrated that the evolution and climax of a baseline rally in tennis could be described in terms of transitions from phasic to non-phasic states. Recently, Bourbonson et al (2010a, 2010b) quantified intra- and inter-team player couplings using relative phase to demonstrate the strength of between-player linkages in basketball. These studies have used lateral and/or longitudinal displacement to calculate relative phase. That approach becomes limited, however, to states of game play where the relative physical orientation of the players is relevant. In larger field-based sports such as soccer, inter-player coordination may not be manifest in movement patterns bound to lateral or longitudinal movement planes. In this paper, we extend on the foundation approach and explore derivatives of player position, and present a new method for quantifying inter-player couplings. Player tracking data recorded using the Amisco® system were compiled from a 45-minute period of an English Premiere League match, and within-team player couplings were calculated using relative phase of player acceleration and angular velocity. Inter-player connections are strongest when temporal synchrony exists in the changes in velocity and/or angle of movement between the playing agents. A method is described that reveals the strongest connections between players within a team, and the relationship to their physical positions on the playing pitch.

Keywords: Relative Phase, Inter-Player Coupling, Soccer, Player Tracking

1. INTRODUCTION

Previous studies have represented the coordinated patterns of movement between limbs in the framework of dynamical systems theory (e.g. Kelso, 1995). A key tenet of this work is that oscillations in joint angles throughout gait cycles for rhythmical activities such as walking or running are characterised by specific phase relations. In this context phase portraits can be used show the relationship between properties of a system with their rate of change (often angular velocity by angular displacement), and phase angles provide a discrete measure to quantify current state of a system.

McGarry et al. (1999) further considered dynamical systems in describing phasic relations in the time course of radial distance of players from the "T" on a squash court during rallies. Generally, the "T" is regarded as a strategically important position from which squash players seek to play. This central location allows the dominant player to reach their opponent's shots more easily, and to economise on movement around the court. As a player moves away from the "T" to return a shot, their opponent invariably aims to return to the centre position. McGarry et al. (1999) revealed frequently anti-phase relations of players oscillating between near and far radial distances from the "T" through the ebb and flow of a rally.

Lames (2006) followed by measuring the difference in phase angle between tennis players during baseline rallies. In this approach, instantaneous phase angles are calculated from a phase portrait representing displacement from the midline of the court and rate of change. Lames (2006) argued that this analysis creates a temporal representation of the cyclical lateral movements of players as they each alternate between moving towards the edge of the court to return a baseline shot, and returning to the centre of the court.

Recently, Bourbonson et al., (2010a, 2010b) quantified intra- and inter-team player couplings using relative phase to demonstrate the strength of between-player linkages in basketball. Phase relations were quantified for the lateral and longitudinal movement of all possible player pairs (Bourbonson, 2010a), and for lateral and longitudinal movement of the team centroids (Bourbonson, 2010b). In both studies, strong in-phase behaviour was reported in the longitudinal (basket-to-basket) direction, and some evidence of in-phase behaviour in the lateral direction (side-to-side).

In many team sports, however, coordinative behaviour may not be manifest simply in physical position, orientation, or even relative direction of movement. Furthermore, Lames (2006) concedes that relative phase measures of lateral movement in tennis can be misleading because considerable longitudinal (i.e. to and from the net) movement occurs, even in baseline rallies. We propose then, that other movement derivatives may provide further insight to the extent of inter-player coordination that are not dependent on comparisons limited to a single the plane or direction of movement.

In this paper we quantify phase relations between player couplings in the domains of acceleration and also angular velocity. Furthermore, we demonstrate that these measures can be used to cluster between-player couplings within a soccer team and discriminate between couplings where movement patterns are the most similar, and those where movement patterns are the least similar. It is proposed that this approach may be helpful in understanding the features of coordinative behaviour between players in team sports.

2. METHODS

Player position data were recorded using the Amisco® video player tracking system, and data were compiled from an English Premier League soccer match. The data were comprised of x-y positions for time at 15 frames per second for a total duration of 45 minutes. Eleven players from a single team were selected for intra-team comparisons. Each of the players participated in the entire sample.

The standard approach to quantifying phase relations is to measure a performance attribute, and plot the normalised attribute with its rate of change. In the context of dynamical systems this representation is often referred to as a phase portrait, and the current state of a system at any moment is measured by the phase angle, which is derived from the slope of a line from the origin to the specified point. Relative phase values were calculated by comparing the phase angle between two players. The calculations of the relative phase values are described below.

Relative phase calculations for acceleration

Velocities for each player were derived from the position data, and normalised using a unity-based normalization

$$V_i = \frac{V - V_{min}}{V_{max} - V_{min}} \quad (1)$$

where V is instantaneous velocity, and V_{min} and V_{max} are the velocity minima and maxima respectively. Acceleration values were derived from the normalized velocities.

$$A_i = \frac{V_i - V_{i-1}}{\Delta t} \quad (2)$$

The phase angle (ϕ_i) for player acceleration was found by

$$\phi_i = \tan^{-1}\left[\frac{A_i}{V_i}\right] \quad (3)$$

Instantaneous phase angles were calculated at each time point for each of the eleven players. Next, relative phase values were calculated for all possible intra-team player couplings (n = 55) by subtracting the acceleration phase angle of a player ($\phi_{player\,A}$) from the acceleration phase angle of his couple ($\phi_{player\,B}$) as follows

$$\epsilon_{AB} = \phi_{player\,A} - \phi_{player\,B} \quad (4)$$

The distributions of relative phase values for all player couplings are shown in Figure 1 below.

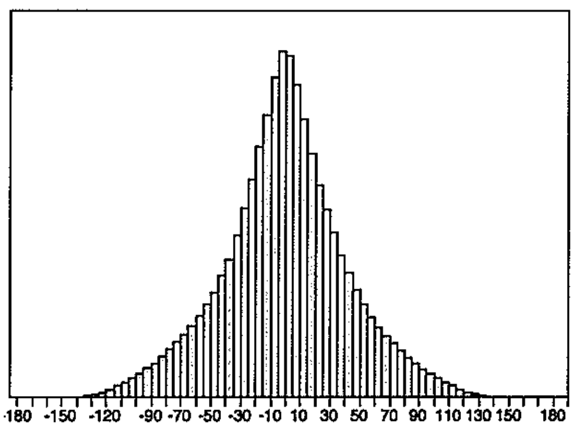


Figure 1: Relative phase distributions for acceleration

Where relative phase is near zero, the acceleration of both players in a coupling could be said to be in-phase. Further, where relative phase is near ±180° the acceleration of both players could be said to be anti-phase. We propose that if two players are accelerating by equal measure, or that one player is accelerating by equal but inverse measure as their coupling is decelerating, then each of these states is equally important in the consideration of the coordination of acceleration. Therefore, we have folded the tails of the relative phase distributions such that the portion of values greater than 90° (or less than -90°) are reversed. As such, a relative phase value of 135° therefore becomes equal to 45°. In this way, in-phase and anti-phase states have equal value.

Relative phase calculations for angular velocity

The angular displacement (θ_j) was calculated from the dot product of consecutive 1-second movement vectors, ***a*** and ***b***

$$\theta_j = \cos^{-1}\left[\frac{\boldsymbol{a}\cdot\boldsymbol{b}}{\|\boldsymbol{a}\|\cdot\|\boldsymbol{b}\|}\right] \quad (5)$$

Next, angular velocity (rate of change in angular displacement) was calculated as follows.

$$\omega_j = \frac{\theta_j - \theta_{j-1}}{\Delta t} \quad (6)$$

The phase angle ($\phi_{\omega j}$) for player angular velocity was therefore found as such.

$$\phi_{\omega j} = \tan^{-1}\left(\frac{\omega_j}{\theta_j}\right) \quad (7)$$

Relative phase values relating to angular velocity ($\epsilon_{\omega AB}$), where therefore calculated using the same method as described in Equation 4.

The distributions of angular velocity relative phase values for all possible player couplings are shown in Figure 2 below.

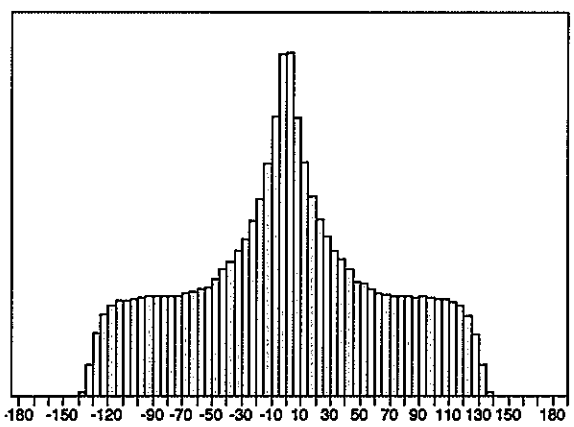


Figure 2: Relative phase distributions for angular velocity

The relative phase distributions for angular velocity were also reversed for values greater than (or less than) 90° using the same method described above for acceleration relative phase distributions. We subsequently refer to these modified distributions as 90°-reversed.

Cluster analysis

A hierarchical clustering algorithm was employed using JMP10 statistical software. This analysis results in the arbitrary grouping of each player coupling with respect to the Euclidean distance between the parameter values. Cluster analysis has been used previously in other sport-related problems to group performance dimensions by similarity (e.g. Chen et al., 2007). In this analysis we added the standard deviations of the 90°-reversed acceleration and angular velocity relative phase distributions for all of the 55 possible player coupling as factors to the cluster analysis using Ward's minimum variance method. Median standard deviations for acceleration and angular velocity relative phase distributions were also calculated for each subsequent cluster group. These figures are presented in Table 2.

Visualization

Finally, the position centroids for each player were calculated using the median position coordinates. These centroids were plotted on a proportional representation of a soccer pitch. The player couplings grouped into two exemplar clusters were overlayed on the pitch to visually illustrate relationship between median pitch position and the degree of movement coordination with nearby (or distant) players.

The nomenclature of soccer-specific player positions was determined by the locale of the position centroids, and the abbreviations for these positions that are used in the results are presented in Table 1.

Position	Abb.
Centre Forward	CF
Left Wide Forward	LWF
Right Wide Forward	RWF
Attacking Midfield	AM
Left Holding Midfield	LHM
Right Holding Midfield	RHM
Left Centre Back	LCB
Right Centre Back	RCB
Left Back	LB
Right Back	RB
Goalkeeper	GK

Table 1: Position abbreviations

3. RESULTS

The cluster analysis identified six discrete clusters where the standard deviations for the 90°-reversed distributions of all possible player couplings for both acceleration and angular velocity relative phase angles were included as factors. The between-player couplings are presented using the position abbreviations, and ordered into the cluster dendrogram presented in Figure 3. The adjacent shaded columns represent the individual coupling means for the angular velocity and acceleration standard deviations respectively. The dendrogram elements are further colour coded by cluster number.

The angular velocity and acceleration relative phase standard deviations for individual player couplings are shown in Figure 4, where each coupling is colour coded by its cluster identity. Additionally, the mean Euclidian distance between each player in the coupling represented in each cluster are shown in Table 2, with the median standard deviations for acceleration and angular velocity by cluster.

Property	Cluster					
	1	2	3	4	5	6
Acceleration Relative Phase	36.58	37.58	38.68	36.53	41.28	39.47
Angular Velocity Relative Phase	50.53	51.35	50.83	50.23	51.20	51.34
Mean Distance (m)	13.1	16.3	21.9	9.8	33.5	29.4

Table 2: Median coupling standard deviations for relative phase angles and mean distance between couplings by cluster.

Cluster 4 was characterised by the lowest median variability in acceleration relative phase, and also the lowest median variability in angular velocity relative phase. This group of player couplings was labelled the "High Coordination Pairs" (HCP) cluster, and couplings grouped in that cluster were inferred to represent between-player movement coordination characterised by a high degree of relative synchrony. Similarly, the couplings in the cluster with the highest median standard deviation for acceleration relative phase, and a relatively high median standard deviation for angular velocity (Cluster 5) was labelled the "Low Coordination Pairs" (LCP) cluster. Couplings in that cluster were inferred to represent between-player movement characterised by a low relative degree of coordination. It can be noted that two other clusters showed higher median variance for angular velocity than the LCP, but both of those showed considerably lower variance for the acceleration property.

Additionally, the HCP cluster was characterised by the lowest mean between-player distance of all clusters (9.8 m), indicating that those players whose movement was more closely synchronized were also more proximal to each other. Further, the mean distance between players in the couplings grouped by the LCP cluster (33.5 m) was greater than for any other cluster, indicating also that the least-most synchronised couplings were also the furthest apart.

Figure 5 presents the overall centroid locations for each player, and also the between-player couplings for members of the LCP and HCP clusters respectively. It is possible to interpret from this figure that the couplings in the LCP cluster are comprised mostly of relationships between the goalkeeper, and the most distal players to the goalkeeper (e.g. CF, LWF, RWF). In contrast, the between-player couplings in the HCP cluster are mostly comprised of a dense group of midfield players, to the exclusion of the goalkeeper, and the players positioned to the periphery of the field.

4. DISCUSSION

Earlier work has explored the utility of dynamical systems for understanding coordinative behaviour between elements in some sports, notably soccer (Grehaigne, 1997), squash (McGarry et al., 2002), tennis (Palut & Zanone, 2005; Lames, 2006), and basketball (Bourbonson et al., 2010a, 2010b). While a dynamical systems approach implies that a sport scenario must demonstrate certain properties that are not considered in this paper, we borrow from the methodologies employed by those studies to measure movement coordination between players.

Specifically, we have aimed to quantify the extent to which pairs of players in professional soccer change speed at the same time, and/or change direction at the same time. Our premise is that in large field-based sports such as soccer, the coordination of player movements are unlikely to be constrained specifically to longitudinal or lateral directions. Therefore we derived phase angles from the relationship between velocity and acceleration, and between angular displacement and angular velocity. These phase angles provide a discrete measure of the current state of movement by each player, regardless of the actual direction of movement. We compared the phase angles between all possible couplings of players within a single team, and found that those data could be meaningfully clustered to rank-order the player couplings from the most coordinated movement activity, to the player couplings with the least coordinated movement activity.

Interestingly, the cluster with the most coordinated player couplings, the HCP group, was primarily comprised of linkages between the central and midfield players. Furthermore, the mean physical distance between those players was the smallest compared to any other coupling cluster. By comparison, the cluster that included the least-coordinated player couplings, the LCP group, consisted mostly of couplings between the goalkeeper, and forwards and wing players who were, on average, most distal from the goalkeepers median pitch position.

These two clusters represent the minima and maxima for the aggregated relative phase data. It is intuitive that the inner midfield players *should* be the most closely coupled, and that the movement relationship between the goalkeeper and the right wing for instance, *should* be comparatively weaker. We have not dealt with the intermediate clusters, other than to demonstrate in Figure 4 that the HCP and LCP clusters could be viewed as opposite ends of a continuum of inter-player movement coordination.

Some interesting outliers are worthy of additional consideration, and are evident in Figure 4. The CF-AM coupling exhibited particularly synchronous acceleration behaviour between the centre forward and the attacking midfield player. The centre forward is often considered as the "target" for the attacking team, and the attacking midfielder is often considered the "playmaker" with the task of executing penetrating passes into attack. The strong relationship observed between these positions reinforces the contextual validity of our approach.

Further, the angular velocity relationship between the RHM-AM pair was stronger than for any other pair. This result indicates that the right holding midfielder and the nearby attacking midfielder changed direction in notably closer synchrony than was evident among any other pairs. The acceleration synchrony between this pair was also amongst the closest of all pairs, which suggests that in overall terms, this player coupling is especially close. It should be noted that the proximity between AM player and both CF, and RHM players is also particularly close.

There are two plausible explanations for these patterns of movement coordination. In a soccer match, players move in response to the position of the ball, and the relative positions of their teammates and players from the opposition. We have examined the relationship between teammates, but neither the ball position, or the pull of opponent players is considered. With regard to the ball, we could speculate that a very large portion of the changes in velocity and angle of movement are likely to be related to similar movements of the ball. If the ball is suddenly kicked to the right attacking side of the pitch, it is clear that many players will change direction and accelerate towards the general location of the ball. Since many of the players in a team are likely to respond similarly, the global attraction of the ball could explain a large amount of apparent between-player movement coordination.

Assuming this is the case, the differences in variability that do exist between couplings could be explained by the additional variance we have observed in between-player coordination. Further study is required to measure the attractor features of the ball, and the any effect that proximity to the ball may have on coordinative behaviour between couplings. It may also be possible to employ approaches that subtract the global attraction of the ball from the movement of all players.

Furthermore, the scope of this paper does not permit us to consider the additional influence of opponent players. An attacker in soccer must move in a coordinated way in relation to their teammates, and the ball, and they must also try to catch their opponent defenders off guard. Clearly these complex interactions are only superficially addressed by our study. Nevertheless, our analysis does exhibit contextual validity; that players change velocity and direction in closer synchrony with proximal players than distal players, and that the clustering groups make some intuitive sense in the context of generic soccer team structures. Closer examination of the additional attractor features of the ball, and interactions with opponent players may provide new insights into team coordination.

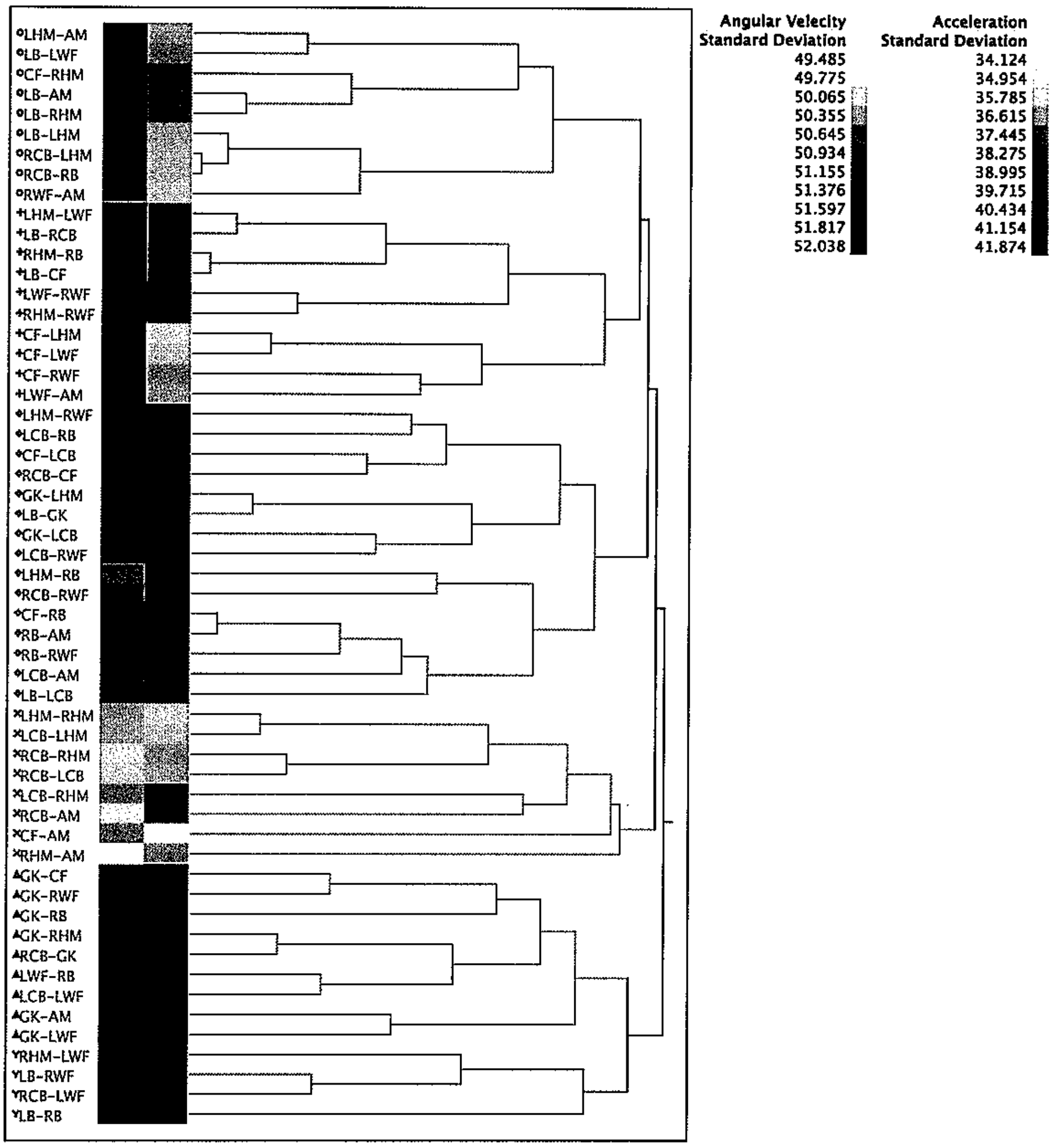


Figure 3: Cluster analysis of player couplings by angular velocity and acceleration relative phase standard deviations.

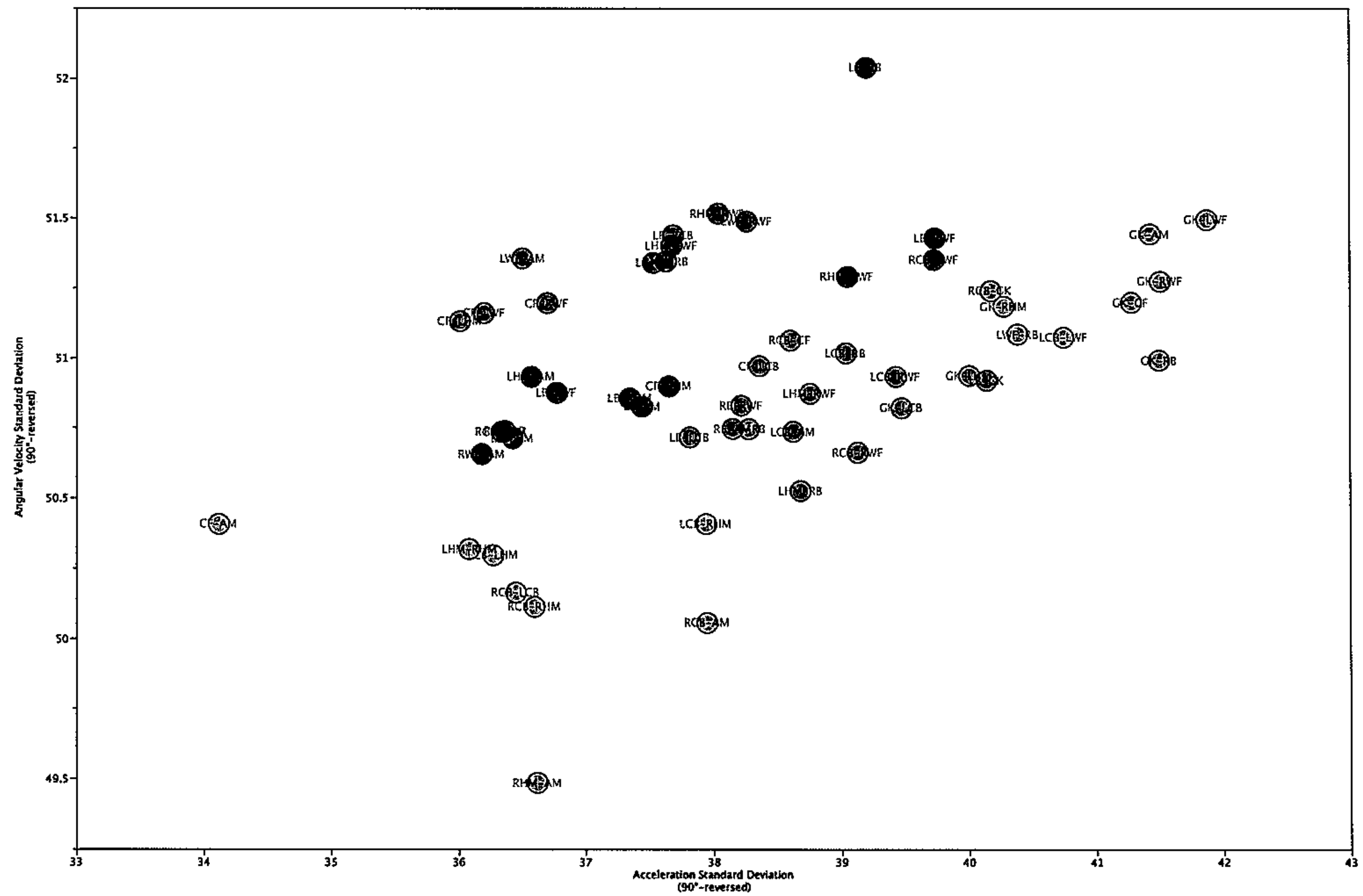


Figure 4: Angular velocity and acceleration relative phase standard deviations with cluster colour.

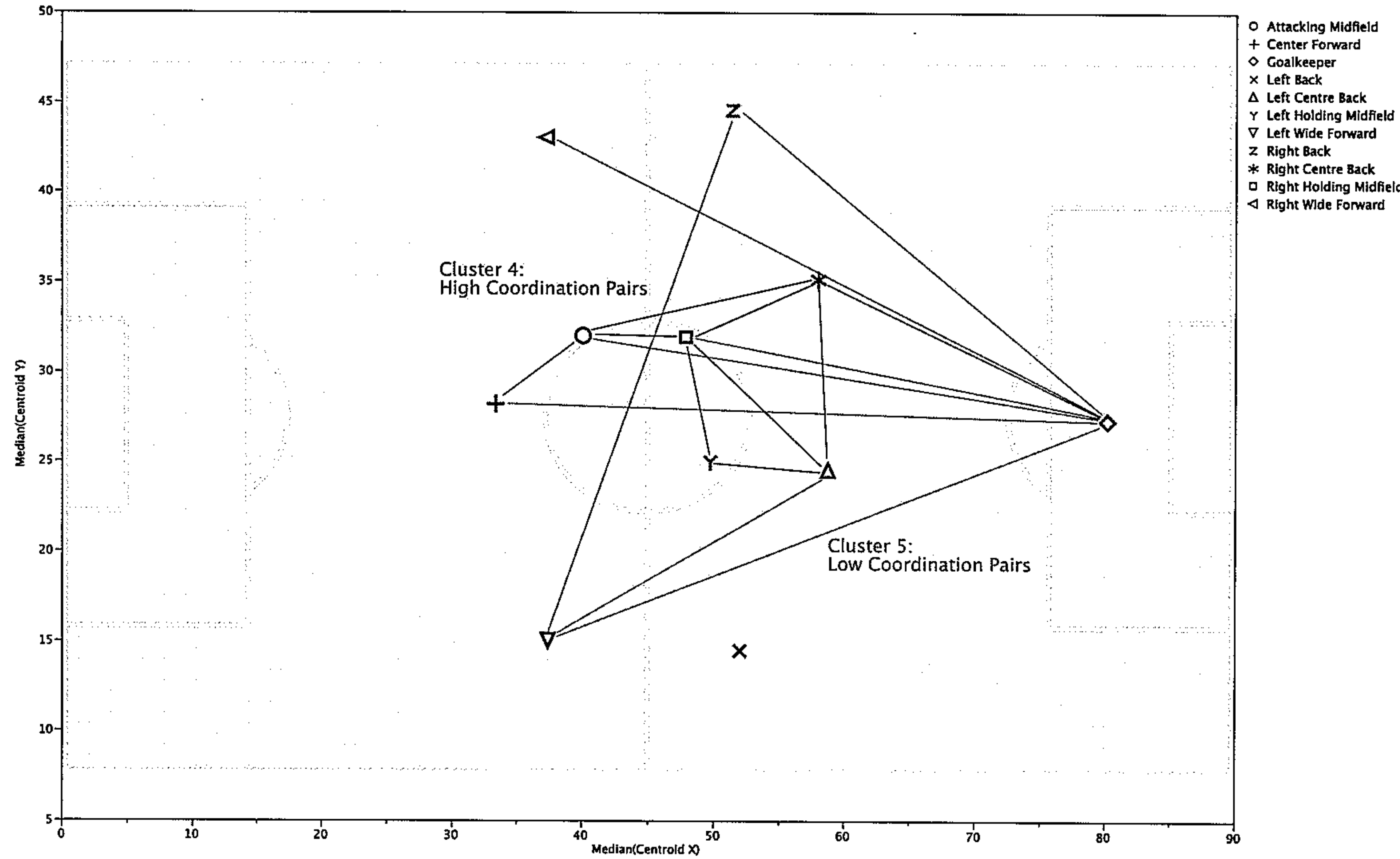


Figure 5: Player position centroids with inter-player couplings for High and Low Coordination Pairs.

5. CONCLUSIONS

Our primary aim is to demonstrate that calculating relative phase angles between couplings, using techniques that are familiar to dynamical systems analysis, provides an insightful representation of the strength of movement coordination between player pairs. We have found that using phase angles to quantify the coordination of changes in velocity, and angular displacement, provide discriminative parameters in a cluster analysis that are consistent with the generic features of soccer game play and position-specific behaviour. Further work may reveal new information about the coordinated behaviour of players in team invasions sports.

References

Bourbousson, J., Seve, C., & McGarry, T. (2010a). Space–time coordination dynamics in basketball: Part 1. Intra- and inter-couplings among player dyads. *Journal of Sports Sciences, 28*, 339-347.

Bourbousson, J., Seve, C., & McGarry, T. (2010b). Space–time coordination dynamics in basketball: Part 2. The interaction between the two teams. *Journal of Sports Sciences, 28*, 349-358.

Chen, I., Homma, H., Jin, C., & Yan, H. (2007). Clustering and display of elite swimmers' race patterns across various comparable criteria at the same time. *International Journal of Sports Science and Engineering, 1*, 129-136.

Grehaigne, J.-F., Bouthier, D., & David, B. (1997). Dynamic-system analysis of opponent relationships in collective actions in soccer. *Journal of Sports Sciences, 15*, 137-149. Kelso, J.A.S. (1995). *Dynamic Patterns: The Self Organization of Brain and Behavior.* Cambridge: MIT Press.

Lames, M. (2006). Modelling the interaction in game sports - relative phase and moving correlations. *Journal of Sports Science and Medicine, 5*, 556-560.

McGarry, T., Khan, M.A. and Franks, I.M. (1999) On the presence and absence of behavioural traits in sport: an example from championship squash match-play. *Journal of Sport Sciences 17*, 297- 311.

McGarry, T., Anderson, D. I., Wallace, S. A., Hughes, M. D., & Franks, I. M. (2002). Sport competition as a dynamical self-organising system. *Journal of Sports Sciences, 20*, 771-781.

Palut, Y., & Zanone, P.G. (2005). A dynamical analysis of tennis: concepts and data. *Journal of Sports Sciences, 23*(10), 1021-1032.

ALLOCATING ENERGY IN A BEST OF 2N-1 MATCH

Pollard, Graham [a, c], Pollard, Geoff [b]
[a] *Faculty of Information Sciences and Engineering, University of Canberra, Australia*
[b] *Faculty of Life and Social Sciences, Swinburne University of Technology, Melbourne, Australia*
[c] *Corresponding author: graham@foulsham.com.au*

Abstract

In this paper we consider a best of 2n-1 games match in which a player can increase his probability of winning a particular game by expending proportionally more of his limited energy within that game. However, a direct consequence of this is that he has less energy available in the very next game and a decreased probability of winning it. Thus, our player has three levels of playing a game (normal, high, low). This characteristic of play is relevant to several sports, and squash is a good example. We consider the cases where he can do this anywhere in the match, as well as when he can do it only a limited number of times. We identify where he should 'swap' two normal games for a high and a low intensity game in order to maximize his probability of winning the match.

This problem has been considered by Brimberg, Hurley and Lior (2004) who considered just a best of 5 (B5) games example. A major objective of the research outlined in this paper was to develop a new structure and method of solution that was simpler from a computational aspect, and one that could much more easily be extended to B2n-1 (n = 4, 5, ...). Another major aim of this study was to link this research to related results in scoring systems research, with the view to identifying and clarifying the underlying reasons why some strategies in this situation are better than others, indeed why some are optimal. Further, we hoped to achieve conclusions that had greater clarity and could be more easily and accurately put into practice.

Keywords: Best of 2n-1 games match; first-to-n match; allocating energy in a match; importance of a win within a point-pair; importance of a draw within a point-pair; fundamental equation of scoring systems.

1. INTRODUCTION

The problem considered in this paper is applicable to several sports, and squash is one example. We assume one player can increase his probability of winning a game by expending proportionally more of his limited energy within that game. We assume that a direct consequence of this is that he has less energy available in the next game, and that his probability of winning that game is lowered. Thus, this player can play two games with probabilities of winning (p, p) or he can play those two games with probabilities (p + δ, p - δ). Our player thus has three levels of playing a game within the best of 2n-1 games match under consideration.

This problem was considered by Brimberg, Hurley and Lior (2004). They gave one best of 5 (B5) example of their methodology with three energy levels, High (H), Medium (M) and Base (B), where the player's probability of winning a game was 0.75, 0.5, and 0.25 respectively. They considered 3 'naïve strategies', (3*H, 2*B), (2*H, 2*M, 1*B) and (1*H, 4*M). It can be seen that these 3 strategies have an H in common. They proved that an optimal dynamic strategy exists by playing all the common elements firstly (in this case playing the common H firstly). They noted that the same principle of playing 'up front' any games that must be played (if all the 2n-1 games were actually played), can be optimally applied at any stage of the match. However, even for their simple B5 example, the solution using their methodology was somewhat complicated. For the B7 case (with correspondingly 4 naïve strategies) their method turns out to be quite cumbersome, and the computational challenges get even worse for B9, B11,.... Their main conclusion was ...'our results on some solved instances of *First-to-2* and *First-to 3* matches suggest that, when a player's back is against the wall (that is he must win all of the remaining games), it is best if he divides his remaining energy evenly over these remaining games. For all other situations, it appears to be best to expend energy with varying degrees of asymmetry between high and low, rather than evenly.'

A major objective of the research outlined in this paper was to develop a simpler structure and method of solution, one that could be more easily extended to B2n-1 (n = 4, 5, ...). A second major aim was to link this research to other work on scoring systems, with the view to identifying the underlying reasons why some strategies were in fact optimal. Further, we hoped to achieve conclusions that had greater clarity and could be more accurately interpreted.

Regarding the related work we make use of the concept of the *importance of a point* within a scoring system. In an elegant paper Morris (1977) defined the importance of a point as the probability that a player wins given he wins that point minus the probability he wins given he loses the point. A useful equation, described here as a fundamental equation of scoring systems, follows from this definition.

A fundamental equation of scoring systems

Suppose I_i is the importance of the i^{th} non-absorbing state within a scoring system when player A has a probability, p, of winning every point. Now suppose player A plans to lift his probability of winning a point to $p + \delta_i$ on a set of points S (if the realization of the game requires him to play some or all of these points in S). Then, the increase in player A's probability of winning under this scoring system, δP, as a result of these increases δ_i, is given by

$$\delta P = \sum_{i \in S} n_i I_i \delta_i \qquad (1)$$

where n_i is the expected number of times state i is visited in one realization of the scoring system when player A's probability of winning a point is equal to $p + \delta_i$ on the set of points S and is equal to p on the remaining points (Pollard, 1986 and 1992). The above equation is exact, whereas the corresponding equation devised by Morris was approximate.

2. METHODS

In this paper we make use of an extension to the above fundamental equation of scoring systems. Insights as to why certain strategies are better than others are achieved through the use of this equation.

2.1 The importance of a point in B2n-1

Suppose player A's probability of winning a point in a B2n-1 game is p. Consider the point being played when player A needs to win a further l points in order to win, whilst player B needs a further m points to win. Then, the importance of that point is given by

$$I(l,m) = \frac{(l+m-2)!}{(l-1)!((m-1)!} p^{l-1}(1-p)^{m-1} \qquad (2)$$

2.2 An extension of the fundamental equation

We give an extension of the above equation from individual points to point-pairs, which can result in a win, a draw or a loss to player A. The importance of winning a point-pair rather than drawing it is equal to the probability of winning given the point-pair is won minus the probability of winning given the point-pair is drawn. Also, the importance of drawing a point-pair rather than losing it is the probability of winning given the point-pair is drawn minus the probability of winning given the point-pair is lost. Thus, using the notation P(i, j) to represent player A's probability of winning from state (i, j) [with player A having won i points and player B having won j points], the importance of winning a point-pair rather than drawing it, $I_{i,j,W}$, is equal to P(i+2, j) - P(i+1, j+1), and the importance of drawing a point-pair rather than losing it, $I_{i,j,D}$, is equal to P(i+1, j+1) - P(i, j+2). The above equation becomes

$$\delta P = \sum n_{i,j}(I_{i,j,W}\delta_{i,j,W} + I_{i,j,D}\delta_{i,j,L}) \qquad (3)$$

where, as before, the importances are evaluated when the p-values are at their 'initial' values, and the $n_{i,j}$ are evaluated when player A has an increase in probability of $\delta_{i,j,W}$ of winning a point-pair at the state (i, j) and a decrease in probability of $\delta_{i,j,L}$ of losing a point-pair at (i, j).

This extension of the fundamental equation from single points with a binary outcome (win/loss) to point-pairs with a tertiary outcome (win/draw/loss) was noted in the paper by Pollard and Pollard (2010). Further, they noted that the equation could be extended to where there were four or more outcomes, and outlined an application to golf.

2.3 B2n-1 games with unlimited 'swapping'

We assume player A has a probability p of winning each game, and q = 1 - p. We also assume that he has the capacity to lift his p-value to p + δ provided he drops it to p - δ on the next game. We call this '*swapping*' (p, p) for (p + δ, p - δ), and in this section we assume that swapping can be done up to n-1 times in a B2n-1 match. The question is where he should choose to do this in order to maximize his probability of winning a B2n-1 match. We use the notation (m, i) to represent the score (or state) in which there is a maximum of m games remaining in the match, and player A is i games ahead of his opponent. The dynamic programming method can be used to determine whether player A should 'swap' or 'not swap' at each game score. Thus, player A's optimal probability of winning given he is in state (2n, i) is given by the general recursive relationship

$$\begin{aligned}P(2n,i) = Max(&p^2P(2n-2,i+2)+2pqP(2n-2,i)+\\&q^2P(2n-2,i-2),(p^2-\delta^2)P(2n-2,i+2)+\\&(2pq+2\delta^2)P(2n-2,i)+\\&(q^2-\delta^2)P(2n-2,i-2))\end{aligned} \quad (4)$$

The optimal decision at each game score (S = swap or N = not swap) and player A's optimal probability of winning from that point score is given in table 1 for all scores that are relevant to the scoring systems B3, B5, B7, B9, B11, and for p = 0.2 (0.1) 0.8 with δ = 0.1. We note that, for all B2n-1 and m = 2, it is optimal to swap if ahead and not to swap if behind. Also, it is clear from the table, and logically, that if one player's optimal strategy at a score is to swap, then his opponent's optimal strategy is not to swap, and vice versa.

State	p = 0.8	p = 0.7	p = 0.6	p = 0.5	p = 0.4	p = 0.3	p = 0.2
(2,1)	S; 0.9700	S; 0.9200	S; 0.8500	S; 0.7600	S; 0.6500	S; 0.5200	S; 0.3700
(2,-1)	N; 0.6400	N; 0.4900	N; 0.3600	N; 0.2500	N; 0.1600	N; 0.0900	N; 0.0400
(4,3)	S; 0.9991	S; 0.9936	S; 0.9775	S; 0.9242	S; 0.8775	S; 0.7696	S; 0.6031
(4,1)	S; 0.9790	S; 0.9240	S; 0.8290	S; 0.6952	S; 0.5310	N; 0.3525	N; 0.1840
(4,-1)	S; 0.8287	S; 0.6572	N; 0.4788	N; 0.3150	N; 0.1808	N; 0.0846	N; 0.0276
(4,-3)	N; 0.4096	N; 0.2401	N; 0.1296	N; 0.0625	N; 0.0256	N; 0.0081	N; 0.0016
(6,5)	S; 1.0000	S; 0.9995	S; 0.9966	S; 0.9862	S; 0.9571	S; 0.8894	S; 0.7500
(6,3)	S; 0.9991	S; 0.9911	S; 0.9631	S; 0.8969	S; 0.7746	S; 0.5878	S; 0.3510
(6,1)	S; 0.9872	S; 0.9361	S; 0.8284	S; 0.6633	S; 0.4604	N; 0.2588	N; 0.1007
(6,-1)	S; 0.9108	S; 0.7519	N; 0.5490	N; 0.3469	N; 0.1810	N; 0.0712	N; 0.0172
(6,-3)	N; 0.6614	N; 0.4229	N; 0.2346	N; 0.1100	N; 0.0412	N; 0.0110	N; 0.0016
(6,-5)	N; 0.2621	N; 0.1176	N; 0.0467	N; 0.0156	N; 0.0041	N; 0.0007	N; 0.0001
(8,3)	S; 0.9993	S; 0.9807	S; 0.9546	S; 0.8623	S; 0.6920	S; 0.4540	N; 0.2067
(8,1)	S; 0.9924	S; 0.9478	S; 0.8337	S; 0.6434	N; 0.4101	N; 0.1965	N; 0.0573
(8,-1)	S; 0.9514	S; 0.8140	S; 0.5996	N; 0.3668	N; 0.1754	N; 0.0586	N; 0.0106
(8,-3)	S; 0.8066	N; 0.5566	N; 0.3177	N; 0.1456	N; 0.0502	N; 0.0114	N; 0.0012
(10,1)	S; 0.9955	S; 0.9577	S; 0.8409	S; 0.6295	N; 0.3707	N; 0.1521	N; 0.0334
(10,-1)	S; 0.9729	S; 0.8576	S; 0.6393	N; 0.3807	N; 0.1679	N; 0.0479	N; 0.0065

Table 1: Optimal swapping strategies for scores relevant to B3, B5, B7, B9 and B11 when δ = 0.1

2.4 B2n-1 with unlimited swapping and p = 0.5

The case of particular interest in sport is when the players are equal, i.e. when p = 0.5 (the middle column of table 1). It can be seen from table 1 that when p = 0.5, player A's optimal strategy is always to swap when ahead and not swap when behind. The case of B7 is given as an example to demonstrate why this is so, and we use the fundamental equation to give an insight into this. Table 2 gives the importance of a win (versus a draw) and the importance of a draw (versus a loss) of the game-pair played at the various game scores when p is fixed at 0.5. It also gives the expected number of times each game-pair score is reached (nij in the earlier equation) in a single play of B7 when the optimal strategy is used. The column δP gives the increase in player A's probability of winning the match resulting from each swap, the sum of these being 0.0051025. Noting from table 1 that player A's optimal probabilities of winning at (6, 1) and (6, -1) are 0.66328 and 0.346925 respectively, player A's optimal probability of winning from the start of the B7 match is thus 0.5051025, giving a total δP of 0.0051025, as in table 2.

It is interesting to note in table 2 that more than half of the increase in player A's probability of winning due to swapping on up to 3 occasions (viz. 0.0051025) is due to the swapping at (2, 1) if it occurs, namely 0.003215.

It can be seen, in this example, that the increased probability of winning a point-pair as a result of swapping is $-\delta^2$ = -0.01, and the decreased probability of losing is δ^2 = 0.01. Thus, *given that the importance of a draw is always greater than the importance of a win when p = 0.5 and player A is ahead, it is always better for player A to swap in this situation. This can be seen to be true for all B2n-1.* Correspondingly, when p = 0.5 and player A is behind, the importance of a win is always greater than the importance of a draw, and so it is optimal for player A not to swap. This is also true when p = 0.5 for all B2n-1.

There is an intuitive or commonsense explanation for the above results. It can be seen that when swapping the outcome is less variable than when not swapping (when swapping the winning probability decreases by δ^2, and the losing probability also decreases by δ^2). *Thus, if player A is ahead, and he wants to stay ahead, it makes good sense to carry out the less variable strategy.* Correspondingly, if player A is behind and he wishes to get ahead, playing the more variable strategy makes sense.

(m,i) Score	Imp^{ce} Win	Imp^{ce} Draw	n(m,i)	δP
(6,1)	0.25	0.375	0.5	0.5(.25*-0.01+.375*0.01)=0.000625
(6,-1)	0.375	0.25	0.5	No swap
(4,3)	0	0.25	0.5*0.6*0.4 = 0.12	0.12*0.25*0.01=0.0003
(4,1)	0.25	0.5	0.5*0.52+0.5*0.25 = 0.385	0.385(0.25*-0.01+0.5*0.01)=0.0009625
(4,-1)	0.5	0.25	0.5*0.24+0.5*0.5 = 0.37	No swap
(4,-3)	0.25	0	0.5*0.25 = 0.125	No swap
(2,1)	0	1	0.12*0.24+0.385*0.52+0.37*0.25=0 3215	0.3215*1*0.01=0.003215
(2,-1)	1	0	0.385*0.24+0.37*0.5+0.125*0.25=.30865	No swap
				Total=0.0051025

Table 2: Analysis of a B7 match when p = 0.5, δ = 0.1, and swapping when ahead (i.e. the optimal strategy)

2.5 B2n-1 with unlimited 'swapping', revisited

The case when player A is better than player B is now considered. For example, suppose p = 0.7 and the players are playing B7. It can be seen from table 1 that player A's optimal strategy when behind in states (6,-1) and (4,-1) is to swap, unlike when the players are equal. This case is analysed in table 3, with the importances being for the case when p = 0.7. (e.g. for state (6, -1), the relevant importances are given by I(3,3) = 0.2646 and I(4, 2) = 0.4116.) Here it can be seen that the increased probability of winning is equal tо 0.0068496. Given that the probability of winning a B7 match when p is fixed at 0.7 is equal to 0.873964 (see table 4), we have that the optimal probability of winning a B7 match when up to 3 swaps are available is equal to 0.873964+0.0068496 = 0.8808136. As a check, noting from table 1 that player A's optimal probabilities of winning from (6, 1) and (6, -1) are 0.936064 and 751896 respectively when p = 0.7, we have that his optimal probability of winning equals 0.8808136, in agreement with above.

Score (m,i)	Imp(win)	Imp(draw)	n(m,i)	δP
(2,1)	0	1	0.316	0.00316
(2,-1)	1	0	0.13064	No swap
(4,3)	0	0.09	0.336	0.0003024
(4,1)	0.09	0.42	0.452	0.0014916
(4,-1)	0.42	0.49	0.188	0.0001316
(4,-3)	0.49	0	0.024	No swap
(6,1)	0.0756	0.2646	0.7	0.001323
(6,-1)	0.2646	0.4116	0.3	0.000441
			Total	0.0068496

Table 3: Analysis of a B7 match with optimal swapping by player A when p = 0.7 and δ = 0.1

Again there is an intuitive explanation for swapping when player A (the better player) is a little behind. It is that, although he is behind, he still has a 'reasonably good' probability of winning. That is, although he is behind, he is 'on track' for a win, and there is an advantage in decreasing the point-pair variation.

Correspondingly, if player A is the weaker player and he is ahead (eg, $p = 0.3$, $\delta = 0.1$ and at scores (6, 1) and (4, 1)), it is in his interests not to swap, as although slightly ahead, he is essentially on track for a loss, and so it is better to be 'more variable' with a higher winning probability. Noting that one player's gain is the opponent's loss, these optimal strategies for player A when $p = 0.7$ and $p = 0.3$ are clearly consistent with each other.

As an aside, we note in table 3 that $\Sigma(\delta P)$ is equal to 0.00316 with a maximum of 2 points left, it is equal to 0.00316 + 0.0019256 with a maximum of 4 points left, and it is 0.00316 + 0.0019256+ 0.001764 with a maximum of 6 points left. Noting the decreasing size of the additional components in these 3 values, it follows, for example, that if player A had only 2 swaps available at the outset rather than the 3 he has in this example, it would appear that he should wait until he has just 4 points left before swapping in order to achieve the greatest return for his swaps. This aspect is considered below.

It can be seen from table 4 that the benefit player A receives from swapping is really quite small. For example, when $p = 0.5$ and $\delta = 0.1$, player A's probability of winning increases by just 0.005 for B3, 0.0051 for B5 and 0.0051025 for B7 as a result of optimal swapping. The corresponding increases can be seen to be (just) a little bigger when $p > 0.5$ (unless p is quite close to 1, a case of little practical relevance), and less when $p < 0.5$. *Overall, it could be said that in practice little is to be gained by simply 'swapping'.*

(p, δ)	P(B3)	P(B5)	P(B7)
(0.8, 0.0)	0.896	0.94208	0.966656
(0.8, 0.1)	0.904	0.94894	0.971886
(0.7, 0.0)	0.784	0.83692	0.873964
(0.7, 0.1)	0.791	0.84396	0.880814
(0.6, 0.0)	0.648	0.68256	0.710208
(0.6, 0.1)	0.654	0.68892	0.716667
(0.5, 0.0)	0.5	0.5	0.5
(0.5, 0.1)	0.505	0.5051	0.5051025
(0.4, 0.0)	0.352	0.31744	0.289792
(0.4, 0.1)	0.356	0.32088	0.292738
(0.3, 0.0)	0.216	0.16308	0.126036
(0.3, 0.1)	0.219	0.16497	0.127489
(0.2, 0.0)	0.104	0.05792	0.033344
(0.2, 0.1)	0.106	0.05888	0.033906

Table 4: Optimal probability, P(B2n-1) of player A winning a B2n-1 match with up to n-1 swaps

It is noted that in Table 4 the first game played in the B2n-1 match has a p-value fixed at p, whereas all other games played can be when the p-value is p, p + δ or p - δ depending on the score. If this game where the p-value is fixed at p is deferred until the 3rd game, the values of P in Table 4 are smaller, and even smaller again if the delay is until the 5th game, etcetera. This is related to a comment in the paragraph preceding Table 7.

2.6 An analysis of B2n-1 with limited 'swapping'

We now consider an example where player A has a limited number of swaps available to him at the outset (namely 3). The dynamic programming relationship similar to that above can be written down, and all the results needed for B3, B5, ..., B11 are given in table 5 where L is the 'number of swaps left', E represents *either* 'swap or not swap' and C stands for 'can't swap' (as he has no swaps left).

State, L	0.8	0.7	0.6	0.5	0.4	0.3	0.2
(2,1), 3	S; .9700	S; .9200	S; .8500	S; .7600	S; .6500	S; .5200	S; .3700
(2,1), 2	S; .9700	S; .9200	S; .8500	S; .7600	S; .6500	S; .5200	S; .3700
(2,1), 1	S; .9700	S; .9200	S; .8500	S; .7600	S; .6500	S; .5200	S; .3700
(2,1), 0	C; .9600	C; .9100	C; .8400	C; .7500	C; .6400	C; .5100	C; .3600
(2,-1),3	N; .6400	N; 0.4900	N; 0.3600	N; 0.2500	N; 0.1600	N; .0900	N; .0400
(2,-1),2	N; .6400	N; .4900	N; .3600	N; .2500	N; .1600	N; .0900	N; .0400
(2,-1),1	N; .6400	N; .4900	N; .3600	N; .2500	N; .1600	N; .0900	N; .0400
(2,-1),0	C; .6400	C; .4900	C; .3600	C; .2500	C; .1600	C; .0900	C; .0400
(4,3), 3	S; .9991	S; .9936	S; .9775	S; .9424	S; .8775	S; .7696	S; .6031
(4,3), 2	S; .9991	S; .9936	S; .9775	S; .9424	S; .8775	S; .7696	S; .6031
(4,3), 1	E; .9988	E; .9928	E; .9760	E; .9400	E; .8740	E; .7648	E; .5968
(4,3), 0	C; .9984	C; .9919	C; .9744	C; .9375	C; .8704	C; .7599	C; .5904
(4,1), 3	S; .9790	S; .9240	S; .8290	S; .6952	S; .5310	N; .3525	N; .1840
(4,1), 2	S; .9790	S; .9240	S; .8290	S; .6952	S; .5310	N; .3525	N; .1840
(4,1), 1	N, .9760	N, .9205	N, .8256	N, .6925	N, .5296	N, .3525	N, .1840
(4,1), 0	C; .9728	C; .9163	C; .8208	C; .6875	C; .5248	C; .3483	C; .1808
(4,-1),3	S; .8287	S; .6572	N; .4788	N; .3150	N; .1808	N; .0846	N; .0276
(4,-1),2	S; .8287	S; .6572	N; .4788	N; .3150	N; .1808	N; .0846	N; .0276
(4,-1),1	N; .8256	N; .6566	N; .4788	N; .3150	N; .1808	N; .0846	N; .0276
(4,-1),0	C; .8192	C; .6517	C; .4752	C; .3125	C; .1792	C; .0837	C; .0272
(4,-3),3	N; .4096	N; .2401	N; .1296	N; .0625	N; .0256	N; .0081	N; .0016
(4,-3),2	N; .4096	N; .2401	N; .1296	N; .0625	N; .0256	N; .0081	N; .0016
(4,-3),1	N; .4096	N; .2401	N; .1296	N; .0625	N; .0256	N; .0081	N; .0016
(4,-3),0	C; .4096	C; .2401	C; .1296	C; .0625	C; .0256	C; .0081	C; .0016
(6,5), 3	S;	S;	S;	S;	S;	S;	S;

	1.000	.9995	.9966	.9862	.9571	.8894	.7500
(6,5), 2	E;	E;	E;	E;	E;	E;	E;
	1.000	.9994	.9964	.9856	.9559	.8871	.7460
(6,5), 1	E;	E;	E;	E;	E;	E;	E;
	1.000	.9994	.9962	.9850	.9546	.8848	.7420
(6,5), 0	C;	C;	C;	C;	C;	C;	C;
	.9999	.9993	.9959	.9844	.9533	.8824	.7379
(6,3), 3	S;	S;	S;	S;	S;	S;	S;
	.9991	.9911	.9631	.8969	.7746	.5878	.3510
(6,3), 2	E;	E;	E;	E;	E;	N;	N;
	.9989	.9905	.9618	.8950	.7724	.5860	.3508
(6,3), 1	N;	N;	N;	N;	N;	N;	N;
	.9987	.9898	.9606	.8931	.7702	.5839	.3487
(6,3), 0	C;	C;	C;	C;	C;	C;	C;
	.9984	.9891	.9590	.8906	.7667	.5798	.3446
(6,1), 3	S;	S;	S;	S;	S;	N;	N;
	.9872	.9361	.8284	.6633	.4604	.2588	.1007
(6,1), 2	E;	E;	N;	N;	N;	N;	N;
	.9859	.9341	.8264	.6620	.4604	.2588	.1007
(6,1), 1	N;	N;	N;	N;	N;	N;	N;
	.9846	.9322	.8243	.6600	.4591	.2583	.1004
(6,1), 0	C;	C;	C;	C;	C;	C;	C;
	.9830	.9295	.8208	.6563	.4557	.2557	.0989
(6,-1),3	S;	S;	N;	N;	N;	N;	N;
	.9108	.7519	.5490	.3469	.1810	.0712	.0172
(6,-1),2	N;	N;	N;	N;	N;	N;	N;
	.9081	.7504	.5490	.3469	.1810	.0712	.0172
(6,-1),1	N;	N;	N;	N;	N;	N;	N;
	.9052	.7484	.5478	.3463	.1807	.0712	.0172
(6,-1),0	C;	C;	C;	C;	C;	C;	C;
	.9011	.7443	.5443	.3438	.1792	.0705	.0170
(6,-3),3	N;	N;	N;	N;	N;	N;	N;
	.6614	.4229	.2346	.1100	.0412	.0110	.0016
(6,-3),2	N;	N;	N;	N;	N;	N;	N;
	.6614	.4229	.2346	.1100	.0412	.0110	.0016
(6,-3),1	N;	N;	N;	N;	N;	N;	N;
	.6595	.4226	.2346	.1100	.0412	.0110	.0016
(6,-3),0	C;	C;	C;	C;	C;	C;	C;
	.6554	.4202	.2333	.1094	.0410	.0109	.0016
(6,-5),3	N;	N;	N;	N;	N;	N;	N;
	.2621	.1176	.0467	.0156	.0041	.0007	.0001
(6,-5),2	N;	N;	N;	N;	N;	N;	N;
	0.2621	.1176	.0467	.0156	.0041	.0007	.0001
(6,-5),1	N;	N;	N;	N;	N;	N;	N;
	.2621	.1176	.0467	.0156	.0041	.0007	.0001
(6,-5),0	C;	C;	C;	C;	C;	C;	C;
	.2621	.1176	.0467	.0156	.0041	.0007	.0001
(8,3), 3	E;	E;	E;	E;	E;	N;	N;
	.9992	.9903	.9536	.8608	.6907	.4537	.2067
(8,3), 2	E;	E;	N;	N;	N;	N;	N;
	.9991	.9898	.9526	.8594	.6894	.4529	.2065
(8,3), 1	N;	N;	N;	N;	N;	N;	N;
	.9989	.9893	.9516	.8578	.6877	.4515	.2055
(8,3), 0	C;	C;	C;	C;	C;	C;	C;
	.9988	.9887	.9502	.8555	.6846	.4482	.2031
(8,1), 3	E;	E;	N;	N;	N;	N;	N;
	.9917	.9465	.8322	.6426	.4101	.1965	.0573
(8,1), 2	N;	N;	N;	N;	N;	N;	N;
	.9911	.9452	.8308	.6415	.4097	.1963	.0573
(8,1), 1	N;	N;	N;	N;	N;	N;	N;
	.9904	.9439	.8291	.6398	.4087	.1960	.0571
(8,1), 0	C;	C;	C;	C;	C;	C;	C;
	.9896	.9420	.8263	.6367	.4059	.1941	.0563
(8,-1),3	N;	N;	N;	N;	N;	N;	N;
	.9497	.8125	.5993	.3668	.1754	.0586	.0106
(8,-1),2	N;	N;	N;	N;	N;	N;	N;
	.9480	.8109	.5986	.3665	.1754	.0586	.0106
(8,-1),1	N;	N;	N;	N;	N;	N;	N;
	.9462	.8091	.5972	.3656	.1751	.0586	.0106
(8,-1),0	C;	C;	C;	C;	C;	C;	C;
	.9437	.8059	.5941	.3633	.1737	.0580	.0104
(8,-3),3	N;	N;	N;	N;	N;	N;	N;
	.8051	.5566	.3177	.1456	.0502	.0114	.0012
(8,-3),2	N;	N;	N;	N;	N;	N;	N;
	.8033	.5559	.3177	.1456	.0502	.0114	.0012
(8,-3),1	N;	N;	N;	N;	N;	N;	N;
	.8008	.5548	.3173	.1455	.0502	.0114	.0012
(8,-3),0	C;	C;	C;	C;	C;	C;	C;
	.7969	.5518	.3154	.1445	.0498	.0113	.0012
(10,1),3	N;	N;	N;	N;	N;	N;	N;
	.9948	.9559	.8387	.6282	.3705	.1521	.0334
(10,1),2	N;	N;	N;	N;	N;	N;	N;
	.9945	.9550	.8375	.6272	.3701	.1519	.0333
(10,1),1	N;	N;	N;	N;	N;	N;	N;
	.9940	.9540	.8361	.6258	.3692	.1516	.0333
(10,1),0	C;	C;	C;	C;	C;	C;	C;
	.9936	.9527	.8338	.6230	.3669	.1503	.0328
(10,-	N;	N;	N;	N;	N;	N;	N;
1),3	.9708	.8551	.6381	.3804	.1679	.0479	.0065
(10,-	N;	N;	N;	N;	N;	N;	N;
1),2	.9698	.8538	.6372	.3800	.1678	.0479	.0065
(10,-	N;	N;	N;	N;	N;	N;	N;
1),1	.9687	.8523	.6359	.3791	.1675	.0478	.0065
(10,-	C;	C;	C;	C;	C;	C;	C;
1),0	.9672	.8497	.6331	.3770	.1662	.0473	.0064

Table 5: Optimal strategies for B2n-1 with limited swapping when p =0.8 (-0.1) 0.2 and δ = 0.1

Table 6 is generated from table 5. It shows player A's optimal strategy for several different states when p = 0.5 and δ = 0.1. It can be seen that when player A is ahead (e.g. by say one game or 3 games) and the maximum duration of the match is not very small, it is in his interests to save his limited number of swaps for possible use later in the match. Indeed, it can be seen from table 5 that, for the various p-values, if player A has just one swap remaining, he can leave that swap until the last two games and still have an optimal strategy. Correspondingly, if player A has just two (or three) swaps remaining, he can leave those swaps until the last four (or six) games and still have an optimal strategy.

State	(2,1)	(4,1)	(6,1)	(8,1)	(4,3)	(6,3)	(8,3)
L= 3	S	S	S	N	S	S	E
L= 2	S	S	N	N	S	E	N
L= 1	S	N	N	N	E	N	N

Table 6: Optimal strategies when p = 0.5 and δ = 0.1.

2.7 An analysis of B2n-1 with an initial lift and unlimited 'swapping'

In the above analyses the game-pair swap was from (p, p) to (p + δ, p - δ). All of the above conclusions are the same if the swaps are to (p – δ, p + δ) rather than (p + δ, p - δ) since, for example, the importance of a win (versus a draw) and the importance of a draw (versus a loss) remain the same. That is, reversing the order of the swaps has no effect on the above conclusions. Thus, if we suppose player A starts a B2n-1 match by lifting his very first p-value to p + δ and then (possibly) swapping during the remainder of the match using this reversed order (p – δ, p + δ), he will never increase his p-value on two successive games (a possibly 'alternating' situation). *It would appear that this is a very practical reality for the situation being modelled.* It turns out that the reward the player receives in this situation of having an initial lift and then optimally swapping (Table 7) is substantially greater than that he receives from simply optimal swapping (Table 4). We have verified that this analysis agrees with that of Brimberg et al (2004) [B5, p =0.5 and δ = 0.25].

Thus, for B2n-1 matches with n small to moderate, whilst player A receives only a small benefit from optimally swapping (up to n-1 times), he receives a considerably larger reward by starting with a lift and then optimally swapping. This greater benefit has been achieved by increasing p by δ up to n times but by decreasing it by δ on one less occasion.

We have also shown that if the 'initial lift' by the player is delayed until the 3rd point (rather than being used on the first point), then the player's P-values in Table 7 are slightly smaller, and are slightly smaller again if it is delayed until the 5th point, etcetera.

P	B3	B5	B7
p=0.8	0.937	0.96397	0.979520
p=0.7	0.834	0.87064	0.89923
p=0.6	0.703	0.72394	0.744612
p=0.5	0.556	0.54312	0.536738
p=0.4	0.405	0.3559	0.320683
p=0.3	0.262	0.19176	0.146243
p=0.2	0.139	0.07452	0.042252

Table 7: Analysis of the case of unlimited swapping with an initial lift (i.e. the 'alternating' case) when $\delta = 0.1$.

2.8 B2n-1 with limited 'swapping' and one lift

We begin this section by noting a general result for B2n-1 matches. It is clear that Player A wins a B2n-1 match if and only if he wins a match consisting of simply playing all the possible 2n-1 games, which can be called the 'full' match. Let us suppose player A's probability of winning the i^{th} game is p_i ($i = 1, 2, \ldots, 2n-1$), $q_i = 1-p_i$ and games are independent. Then, using the probability generating function for player A's score minus player B's score for the 'full' match, it can be seen that the probability player A wins the B2n-1 match is equal to the sum of the coefficients of $x^1, x^3, x^5, \ldots, x^{2n-1}$ in the expression $\Pi_i(p_i x^1 + q_i x^{-1})$. Thus, it is clear that player A's probability of winning does not depend on the ordering of $p_1, p_2, \ldots, p_{2n-1}$. This is a slight extension of the paper by Kingston (1976), who considered the case in which the p_is were equal.

For the analysis of the B2n-1 match with limited swapping and one lift, we consider an example with $p = 0.5$ and $\delta = 0.1$. Suppose player A is playing a B11 game and has 2 swaps and one lift available to him at the outset. When is it optimal for him to use those two swaps and the lift? We can see from the section above that the two swaps can be used optimally in the last 4 possible games, that is, on the 8th, 9th, 10th and 11th games if played. It can be seen, as in the previous paragraph, that the lift can be used on the 1st, 2nd, …, or the 7th point with equal effect. Given the lift is used on one and only one of the first seven points, table 8 gives the probability player A wins in 6 or 7 points, the probabilities that the score (m,i) reaches (4,3), (4,1) , (4,-1) and (4,-3) [note that these probabilities, as in the previous paragraph, also do not depend on where in the first 7 points the lift actually occurs], and the probability player A wins from these state given he has two lifts available (see table 5). The optimal probability player A wins a B11 given he has two swaps and one lift available at the outset is thus 0.528311875, in agreement with table 5 [0.6* 0.6272+ 0.4* 0.3800 = 0.5283, to 4 decimal points], which gives the dynamic programming solution. Note that it is optimal for the swaps to be used at the end of the match, and, if the lift occurs before the swaps, there is an additional benefit in the swaps because they occur with higher probability. That is, there is clearly a positive interaction in the lift occurring before the swaps.

			Win Probability	Product
Probability A wins in 6 or 7 points	0.071875			
Probability score reaches (4,3)		0.178125	0.9424	0.167865
Probability score reaches (4,1)		0.28125	0.6952	0.195525
Probability score reaches (4,-1)		0.265625	0.315	0.083671875
Probability score reaches (4,-3)		0.15	0.0625	0.009375
Probability A loses in 6 or 7 points	0.053125			
			Probability A wins	0.528311875

Table 8: Analysis of B11 with one lift and 2 swaps, when $p = 0.5$ and $\delta = 0.1$

We can observe in detail the positive interaction between swapping (say swapping up to twice on the last four games of B11) and lifting prior to the swapping. Given $p = 0.5$, $\delta = 0.1$ and swapping up to twice, the probability player A wins B11 when lifting on one of the first 7 games (if all are played) is equal to 0.528311875 (as in Table 8), and when not lifting on any of those 7 games is equal to

0.5*0.6272 + 0.5*0.3800 = 0.5036 (from table 5, or more accurately 0.503592969). Also, the probability player A wins given he never swaps, but lifts (or plans to lift) on just one game in the B11 match, is equal to 0.6*0.6230 + 0.4*0.3770 = 0.5246, from Table 5 (or more accurately 0.524609375). Thus, the increase in player A's probability of winning due to just swapping (up to twice) is 0.003592969, and the increase due to (up to) one lift is 0.024609375, and these sum to 0.028202344 which is less than 0.028311875 from Table 8. Thus, there is a very small but positive interaction of (0.028311875 – 0.028202344) = 0.000109531 between the two swaps and the lift.

3. DISCUSSION

Brimberg et al (2004) considered only one example, a B5 example. That example involved a player who was better than his opponent. Their assumptions in that example were equivalent to a player who, in the terminology of this paper, could lift on one point as well as swap twice. Lifting and swapping were not separated in their paper. As a result it was not clear the extend to which the benefit to their player was due to the equivalent of lifting and the extent it was due to the equivalent of swapping. Further, it was not clear the extent to which their player's optimal winning probability, being greater than 0.5, was actually due to him being a better player rather than due to him swapping optimally. We believe these issues have been clarified in this paper.

4. CONCLUSIONS

In this paper we have considered the situation where a player has the capacity to increase his probability of winning a game from p to p + δ provided he decreases it by δ on the next game. We called this '*swapping*' the game-pair (p, p) to (p + δ, p - δ). We included the case where the player can also *lift* his p-value by δ on a single game, e.g. the first. This leads naturally to the case in which the player lifts on the first game, and then possibly 'alternates' by swapping in the form of decreasing p and then increasing it, thus never increasing p on two consecutive games. The scores in a B2n-1 match at which the player should choose to swap and lift in order to maximize his probability of winning have been determined.

It has been shown that, for all B2n-1 matches, it is optimal, when the players are equal (i.e. p = 0.5), for a player to swap when ahead, and not to swap when behind. It was noted that the game-pair after swapping has a 'less variable' outcome (than without swapping), and so an 'equal' player who happens to be ahead is more likely to stay ahead by swapping. This intuitive explanation complements the mathematical analysis.

It can sometimes be optimal for the better player to swap even when slightly behind. An explanation is that, although he is slightly behind, he is 'on track' for a win, and so there is an advantage in decreasing the game-pair variation by swapping. Likewise, if a player is the weaker player and he is slightly ahead, it can be optimal for him not to swap, as although he is slightly ahead, he is essentially on track for a loss, and so it is better for him to be 'more variable' in order to increase his chance of winning. Noting that one player's gain is the other player's loss, these optimal strategies for the better and weaker players are clearly consistent with each other.

The benefit a player receives from just optimal swapping has been shown to be very small. However, for B2n-1 matches with a small to moderate value for n, whilst a player receives only a small benefit from swapping on up to n-1 occasions, he receives a considerably larger reward by lifting on the very first point and then using optimal swapping. This substantially greater benefit is achieved by increasing p by δ on up to n occasions but by decreasing it by δ on up to only n-1 occasions. This 'alternating' case is a case of real practical relevance.

The case of a limited number of swaps was also considered. It was shown that when the players are equal and player A is a little ahead (e.g. by say one game or 3 games) and the maximum remaining duration of the match is not very small, it is in his interests to save his limited number of swaps for possible use later in the match. Indeed, it was shown that, for a range of p-values, if player A has just one swap remaining, he can leave that swap until the last two games for possible use, and still have an optimal strategy. Correspondingly, if player A has just two (or three) swaps remaining, he can leave those swaps until the last four (or six) games for possible use, and still have an optimal strategy.

The situation of one lift and a limited number of swaps was also considered. As an example the case of B11, p = 0.5, with player A having 2 swaps and one lift available to him at the outset, was considered. The optimal strategy was shown to be to use the two swaps on the 8^{th}, 9^{th}, 10^{th} and 11^{th} games if appropriate (i.e. on last possible 4 games), and to use the lift on any one of the first 7 games. The lifting on one of the first 7 games and the possible swapping afterwards had a positive interaction.

Concluding, the use of game-pairs rather than individual games has simplified this problem. The importances of game-pairs and the fundamental equation of scoring systems have been used to explain mathematically and intuitively why certain strategies are best. As a result the conclusions can be more easily interpreted by players.

References

Brimberg, Hurley and Lior (2004). Allocating energy in a first-to-n match. IMA journal of Management Mathematics, 15, 25-37.

Kingston, J. G. (1976). Comparisons of scoring systems in two-sided competitions. Journal of Combinatorial Theory, Series A, 20, 357-362.

Morris, C. (1977). The most important points in tennis. In Optimal Strategies in Sports, edited by S. P. Ladany and R. E. Machol, 131-140. Amsterdam:North-Holland. (Vol 5 in Studies in Management Science and Systems).

Pollard, G. H. (1986). A Stochastic Analysis of Scoring Systems. PhD Thesis. Australian National University.

Pollard, G. H. (1992). The optimal test for selecting the greater of two binomial probabilities, The Australian Journal of Statistics, 34, 273-284.

Pollard, G. H. & Pollard. G. N. (2010). The efficiency of scoring systems with outcome dependent points. Proceedings of the Tenth Australasian Conference on Mathematics and Computers in Sport, edited by A. Bedford and M. Ovens, Mathsport (ANZIAM), 35-42.

IN-PLAY MODIFICATION OF SERVICE PROBABILITY IN TENNIS

Michelle Viney, Elsuida Kondo and Anthony Bedford

School of Mathematical and Geospatial Sciences
RMIT University, Melbourne, Australia
Corresponding author: anthony.bedford@rmit.edu.au

Abstract

In this paper, a set simulator for tennis was built to look at the variation in set outcomes, as the service probability is updated. A simulation was compared to establish Markov results, and to determine its validity. Three methods were developed to update the server's probability - when a player wins serves after a game, during a game, and by court-side. Comparing the three methods using simulator data and one case study, the results found the three methods to be quite close in comparison to the Markov model.

Keywords: In-play, Markov, service probability, tennis.

1. INTRODUCTION

For many years Markov models, and their associated results in tennis, have been at the forefront of academic research. One of the great advantages introduced by online and mobile technology in the last decade has been the advent of multiple in-play wagering markets. These markets allow trading to take place during the actual event (such as a set winner during the set), or close to the commencement of an event (such as the new game winner).
Notably, a hierarchical Markov model is most often used to calculate the probability of a player winning. The hierarchical scoring system of the game is well served by such a model: a Markov server-dependent model, where points are required to win games, games required to win sets, and sets required to win the match. Given that the probability of a player winning a point on his serve is assumed to be constant, a Markov chain can be assembled with different states demonstrating different scorelines.

Notably, some research suggests the possible need to update the probability of winning on serve whilst in-play (Klaassen and Magnus 2003). The following questions are raised: (1) by what amount should the probabilities increase/decrease in-play, (2) should the serve probability remain fixed or vary based upon the stage of the game or set and, (3) what happens when the player performances deteriorates or the other player starts to improve, i.e. should that player remain at the current probability or should it be decreased? We considered answering these questions through a comparative simulation method, braced against a Markov and semi-Markov model for comparison.

2. METHODS

Figure 1 outlines a Markov chain demonstrating the outcome of a particular game in which a server has a probability, p, of winning a point on his serve.
A well known fact is that tennis models assume the probability of winning a point on service is constant, and thus a model can be produced, given the assumption of points within a match being identically and independently distributed. Klaassen and Magnus (2001) examined this theory, and as a result, they determined that winning the previous point has a positive effect on winning the current point.

Klaassen and Magnus (2003) also describe a method to forecast a winner of a tennis match not only at the beginning of the match, but during the match, with the focus on ATP rankings and point-by-point data. In their research they assumed that points were IID and the input probabilities are *not* reorganised as the match is in progress. In conclusion they state "One could think of a Bayesian updating rule, where the prior estimates are $\hat{p}_a$ and $\hat{p}_b$, obtained before the match starts, and the likelihood comprises the match information up to the current point."

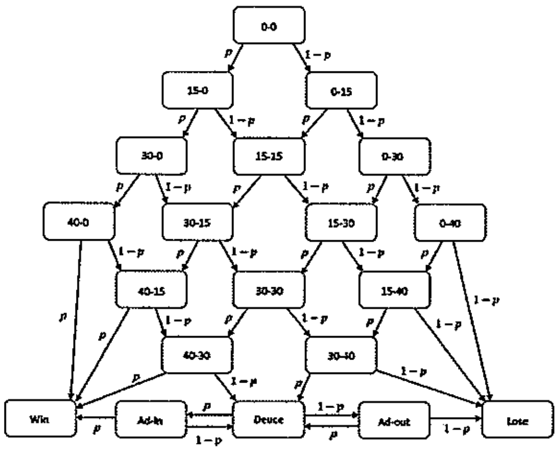


Figure 1: A Markov Chain of a tennis game

Using the Markov chain model, Barnett (2006) demonstrated that updating the prior estimates as the match is in-progress can improve predictions, showing updating is superior to steady values.
The Markov chain model is typically still used to predict outcomes of tennis matches before and during the match. Barnett, Brown & Clarke (2006) used the properties of the Markov Chain to derive a recursive formula to calculate the probability of winning from any state within a game, set and match. In terms of a game, the probability of player A winning the game at point score (a, b) is given by:

$$P(a,b) = pP(a+1,b) + (1-p)P(a,b+1) \quad (1)$$

with boundary conditions:

$P(a,b) = 1$ if $a = 4, b \leq 2$
$P(a,b) = 0$ if $b = 4, a \leq 2$
$P(3,3) = \frac{p^2}{p^2+(1-p)^2}$,
where p is the probability of player A winning a point on serve and remains constant for the entire match.

In a similar fashion, the probability of either player winning a tiebreak set can be calculated using a Markov chain. Let $P_A^{gs_T}(c,d)$ represent the conditional probability of player A winning a tiebreak set from game score (c, d) where player A is serving. It is expressed as followed:

$$P_A^{gs_T}(c,d) = p_A^g P_B^{gs_T}(c+1,d) + (1-p_A^g)P_B^{gs_T}(c,d+1) \quad (2)$$

with boundary conditions

$P_A^{gs_T}(c,d) = 1$ if $c = 6, 0 \leq d \leq 4, c = 7, d = 5$
$P_A^{gs_T}(c,d) = 0$ if $d = 6, 0 \leq c \leq 4, c = 5, d = 7$
$P_A^{gs_T}(6,6) = p_A^{gT}$,

where p_A^g represents the probability of player A winning a game on serve and p_A^{gT} represent the probability of player A winning a tiebreak game. For a detailed explanation, see Barnett, Brown & Clarke (2006).

To determine whether updating the probability is more or less effective compared to the Markov Chain Model, three different methods were produced. They are *Post-game*, *In-game* and *Court-side*. *Post-game* involves updating the probability of the server winning a point on serve after the game is complete. It involves multiplying the proportion of points won by the server in the game by a theta value and then adding or subtracting (dependent on the outcome of the game) to the probability of winning a serve from the previous service game. For example, Player A has lost their service game. Then the proportion of points Player A won whilst serving is multiplied by a chosen weighting parameter. Then this value is decreased from Player A's probability of winning a point on serve. This updated probability is then implemented the next time Player A serves.
The *Post-game* equation for Player A serving at game, g is

$$p^*_{a,g[pg]} = p^*_{a,g-1,[pg]} + 1_{\{A\ wins\}} \cdot - 1_{\{A\ loses\}} \frac{x}{\theta n} \quad (3)$$

If $x = 0$, then

$$p^*_{a,g[pg]} = p^*_{a,g-1[pg]} + 1_{\{A\ wins\}} \cdot - 1_{\{A\ loses\}} \frac{1}{\theta} \quad (4)$$

where $\theta =$ the weighting parameter; $g =$ game number; $n =$ the number of points in a game, g; $x =$ the number of points Player A has won in the game; $p^*_{a,0[pg]} = p_a$ and as $\theta \to \infty, p^*_{a,g[pg]} = p^*_{a,g-1[pg]}$. The receiver Player B probability does not alter, therefore $p^*_{b,g[pg]} = p^*_{b,g-1[pg]}$.

In-game involves updating the server's probability of serving after each point is complete. For example if the server lost the previous service point, then their probability is decreased by the chosen weighting parameter. The *In-game* equation for Player A serving at point, p is

$$p^*_{a,p[ig]} = p^*_{a,p-1[ig]} + 1_{\{A\,wins\}}, -1_{\{A\,loses\}}\frac{1}{\theta} \quad (5)$$

where $\theta =$ the weighting parameter, $p =$ point number, $p^*_{a,0} = p_a$ and $p^*_{b,p} = p^*_{b,p-1}$.

The *Court-side* approach takes into consideration the side of the court the server has won or lost the point, and updating the probability of serve for the next time the server serves on that particular side of the court. In tennis they are two sides of the court to serve from, deuce and advantage court. For example if the server lost the point on the advantage court, then the next time the server serves on the advantage court, their probability is decreased by a weighting parameter.

For Player A serving at point, *p*

$$p^*_{a,p[cs]} = p^*_{a,p-2[cs]} + 1_{\{A\,wins\}}, -1_{\{A\,loses\}}\frac{1}{\theta} \quad (6)$$

where $\theta =$ the weighting parameter, $p^*_{a,0[cs]} = p_a$ and $p^*_{b,p[cs]} = p^*_{b,p-2[cs]}$.

In all three methods the updated probabilities are carried over to be implemented when Player A serves next.

The Markov Chain method, involves assuming all points are independent and identically distributed (IID), therefore the probability of winning a point on serve remains fixed for the match.

2.1 Case study

Consider a match between Rafael Nadal and Andy Roddick, Nadal is serving at the start of the set with a serve probability of 0.60 and Roddick's serve probability of 0.62. In this case study all methods have a theta value of 200. Table 1 shows the amount of variability Nadal's probability of winning a point on serve will vary depending on which method is selected. For the Post-game method, Nadal wins four of out six points, therefore Nadal's probability of winning a point on serve increases by 0.003. For the In-game approach, as Nadal wins or loses a point, his probability changes after every point.

As an example, for the first point in the game, Nadal has won the point therefore his probability increases to 0.605. In terms of Court-side, Nadal won the first point on the deuce side, thus his probability increases and is applied to the next time Nadal serves on the deuce side. It is important to note that the receiver's probability does not update.

2.2 Implementing

To compare the four approaches, a set simulator was built using the add-on @Risk for Microsoft Excel. This simulator has the ability to modify the probabilities as the simulation is occurring, and the set can commence at any set score for any server. Once a simulation is complete the output displays the number of times each player has won to a particular set score.

Before the simulation can occur, the following input parameters were entered into the simulator: the probability of winning on serve for both players, the game score and the server of the current game. To determine the winner of the point, a uniform distributed random number is generated in order to compare the server's probabilities. This process of generating a uniformly random number is repeated many times until the set is complete.

In order to compare the different models, all models were linked to each other to ensure all methods have the same random variable value.

3. RESULTS

The simulation was first performed for two scenarios with different probabilities of winning a point on serve. The first scenario is a "Balanced situation" where both players have a probability of winning a point on serve at 0.5. The second scenario reflects a more realistic match where Player 1's probability of winning a point on serve is 0.60 and Player 2 is 0.62.In both scenarios, Player 1 serves first in the set. The simulator was run for the two scenarios with five different weighting parameters.

To compare the three methods to the Markov Chain model, a Pearson chi square test was performed to determine whether any statistically significant difference occurred on the outcomes.

The three proposed methods cannot have the same theta values, as the *Post-game* method only updates the probability of the server after the game is complete, therefore having a large theta value will result in a minimal updated probability. The *In-game* and *Court-side* method have the same theta value for all simulations: the probabilities are updated after every point for both methods.

	Post-game	In-game	Court-side	Markov	Score	Side of court	Wins
Game 1							
	0.600	0.600	0.600	0.600	0-0	Deuce	Y
	...	0.605	0.600	...	15-0	Advantage	N
	...	0.600	0.605	...	15-15	Deuce	Y
	...	0.605	0.595	...	30-15	Advantage	N
	...	0.600	0.610	...	30-30	Deuce	Y
	...	0.605	0.590	...	40-30	Advantage	Y
Game 3							
	0.603	0.610	0.615	0.600		Deuce	
			0.595			Advantage	

Table 1: A representation on Nadal's probability of winning a point on serve changing within the game with weighting parameter 200

Sim Number	Post-game	Court-side	In-game
1	60	200	200
2	80	250	250
3	100	300	300
4	120	400	400
5	200	500	500

Table 2: Theta values used in the simulation

Table 2 outlines the theta values that were applied to the simulator. The value of theta does not have to be fixed, but in this paper, the theta values remain fixed.

The simulation was performed on all theta values with 10,000 simulations. The key was to determine, using a balanced model, which values were close to the Markov results. Our analysis found that in both scenarios, *Court-side* with simulation case four and five and *Post-game* simulation case five, were not statistically significant at a 0.05 level.
In both scenarios all the cases for the *In-game* approach were statistically significant at a 0.05 level. The results suggest that the *In-game* method was the most diverse (to the Markov model) compared to the other proposed methods. To validate this statement further analysis is required on empirical data.

Once we decided the values of theta, we needed to determine if 10,000 simulations was enough to provide valid results. To determine the optimal value, the Markov Chain simulator was run four times (with different simulation lengths) using a balanced match with Player 1 serving first. This was then compared against the theoretical Markov model results. The four simulation lengths were 10,000, 20,000, 50,000 and 100,000 simulations.

Table 3 outlines the amount of times each player wins a particular set score by simulation length. For example, Player 1 has approximately a ten precent chance of winning the set to the set score 6-3. All Markov simulator quantities are very close to the theoretical Markov Model. This also shows that as the quantities increased from 10,000 to 100,000 simulations, the error rate decreases from 0.24% to 0.14%.

Table 3 also shows that in the theoretical Markov model, both players have the same probability of winning at all set scores, as expected. It can be noted that the theoretical Markov model does not take into account any real in-game advantage. For example, to win the set score of 6-1, the server who serves first in the match will serve four times compared to the other player who only serve three times. For the set

Markov Simulation	Wins 6-0		Wins 6-1		Wins 6-2		Wins 6-3		Wins 6-4		Wins 7-5		Wins 7-6	
	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2
10000	0.015	0.017	0.049	0.050	0.083	0.083	0.112	0.108	0.125	0.126	0.057	0.057	0.058	0.060
20000	0.017	0.017	0.053	0.048	0.080	0.081	0.108	0.107	0.124	0.122	0.059	0.060	0.061	0.064
50000	0.017	0.017	0.051	0.049	0.083	0.082	0.108	0.107	0.121	0.124	0.062	0.059	0.060	0.062
100000	0.018	0.018	0.048	0.049	0.084	0.082	0.110	0.108	0.119	0.121	0.061	0.061	0.062	0.060
Markov	0.016	0.016	0.047	0.047	0.082	0.082	0.109	0.109	0.123	0.123	0.062	0.062	0.062	0.062

Table 3: Comparing different simulations quantities with the theoretical Markov Chain model

scores 6-1 and 6-3, it could be assumed that the server, who serves first in the set, has an advantage if there is the presence of a cumulative advantage and therefore should have a higher proportion of times of winning at those set scores than their opponent. This question should be further explored on empirical data.

4. DISCUSSION

A case study was performed where Jesse Levine and Ryan Harrison who played at the Newport ATP tournament in July, 2012. The starting set price by Bet365 was $2.00 for Levine and $1.75 for Harrison. The probability of winning a point on serve was estimated with Levine at 0.6005 and Harrison 0.6095. Twenty thousand simulations were performed for all weighting parameters on all four methods. A Pearson chi square test was performed at a five percent significance level and it was concluded that all combinations were statistically significant with the exception of one case which was *Court-side* with weighting parameter 500.

For further analysis, case three of the theta values were selected (weighting parameter 100 for *Post-game* and 300 for *Court-side* and *In-game*). The set simulator was then performed of all four methods with the desired weighting parameters. Figure 1 displays the amount of times Levine or Harrison may win to a particular set score over 20,000 simulations. For example the least likely set score for each player winning is 6-0. It shows that Harrison has a higher number of wins at 6-0 because his starting probability is greater than Levine. Overall the graph shows that all three models reflect the Markov model quite closely. In the actual match, the first set was won by Harrison 6-3. Looking at Figure 1, the 6-3 had the highest probability. Numerically speaking, the theoretical Markov estimated Harrison to win 18% at 6-3 whereas the other models were estimating around 17%.

In terms of selecting the optimal method in updating probabilities, no one method can be chosen as superior as yet. Future research is required to determine the optimal theta value and method in light of real data. Although the preferred method has not been determined, in context to the game of tennis, the *Court-side* method may best reflect the true match. This is due to some players being more dominate on a certain side. For example the left handed serve on the backhand side is considered a strength. The *Court-side* approach could be taken one step further by looking into the affects of ends

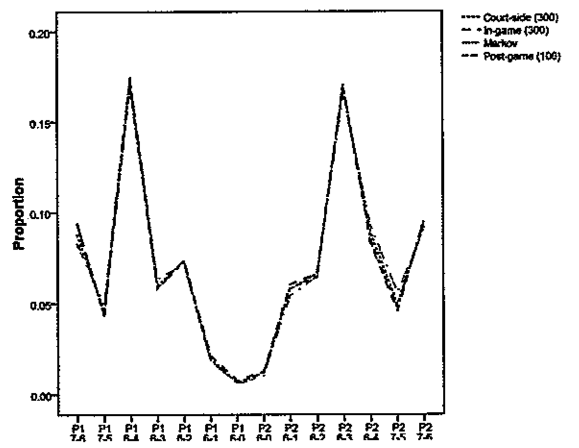


Figure 1: Representing the amount of times each player may win to a particular set score for 20000 simulations. P1= Levine and P2= Harrison

and how the sun or wind could affect the game. Therefore this approach would involve updating the probability on the server only when they are on the same court side and end.

For future research we shall update the probabilities after every point and to simulate to compare against the wagering markets. In this case study, the analysis was performed where Harrison was leading at 3-1 with Harrison to serve next. To reflect the state of the match, the probabilities were updated on the three methods. The simulator was run again on case three with 20,000 simulations. Table 4 displays the wagering odds and the simulated results, the theoretical Markov odds and Bet365 odds. Table 4 shows that a comparison of all three proposed methods seems to be similar. When comparing the simulated results with the wagering odds offered by Bet365, it was noted that in some instances Bet365 has a higher probability as compared to the simulated results (e.g. for Harrison to win at 6-1, 6-2 and 6-3). But it is important to note that Bet365 has to take into account its margin, therefore in this case, the over-round may have gone to those set scores. The actual result of the set was Harrison to win the set at 6-3. In terms of these results, all methods had Harrison at the highest probability to win at 6-3.

Another feature to explore was looking into the number of serves won on serve and return for both players. To analyse this, a Markov Chain Model with 20,000 simulations was performed.

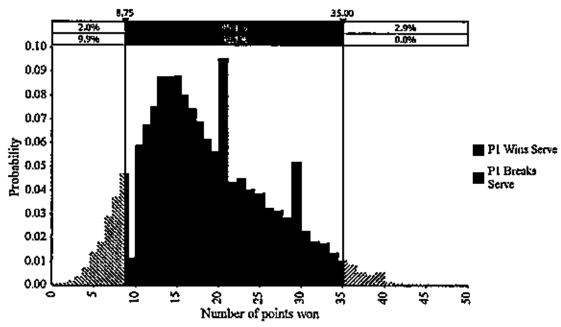


Figure 2: The distribution of Player 1 (Levine) winning points on serve and returning

mean score both on serve and returning compare to Levine. When comparing the two graphs, it is interesting to observe that Harrison's two distributions are further away from Levine's. This could be an affect of having a higher probability on serve, and therefore a higher likelihood of holding serve. This is an interesting aspect to conduct further research in.

Overall the three proposed models performed well against the Markov chain model. For further research it is recommended to perform simulations at a quantity greater than 10,000, given the improved convergence of these probabilities, and the likely unbalanced values for the service probability under the proposed changes. All the results are based on simulated data and one case study, therefore all these findings need to be validated against empirical data, before a conclusion can be made.

Method	Wins 6-1		Wins 6-2		Wins 6-3		Wins 6-4		Wins 7-5		Wins 7-6	
	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2
Court-side	-	6.09	-	7.20	108.11	2.83	18.66	12.97	32.47	24.57	15.27	14.98
In-game	-	5.84	-	7.09	117.65	2.71	20.02	14.57	35.52	28.13	16.52	14.84
Post-game	-	6.25	-	7.53	126.58	2.65	19.74	14.37	32.84	28.53	15.33	14.12
Markov	-	6.59	-	7.44	106.38	2.94	16.81	11.95	26.99	22.96	14.86	13.71
Theoretical Markov	-	6.66	-	7.34	125.57	2.90	17.10	13.51	24.32	22.07	14.56	13.73
Bet365	-	5.00	-	6.00	101.00	2.25	17.00	15.00	23.00	21.00	13.00	12.00

Table 4: Wagering odds offered in all methods at set score 3-1, Player 2 (Harrison leading and serving next)

Figure 2 outlines the distribution of the number of points Levine has won on serve and point won returning. Levine's mean number of points winning a point on serve for the set is 21.58 compared to winning a point whilst returning is 15.04.

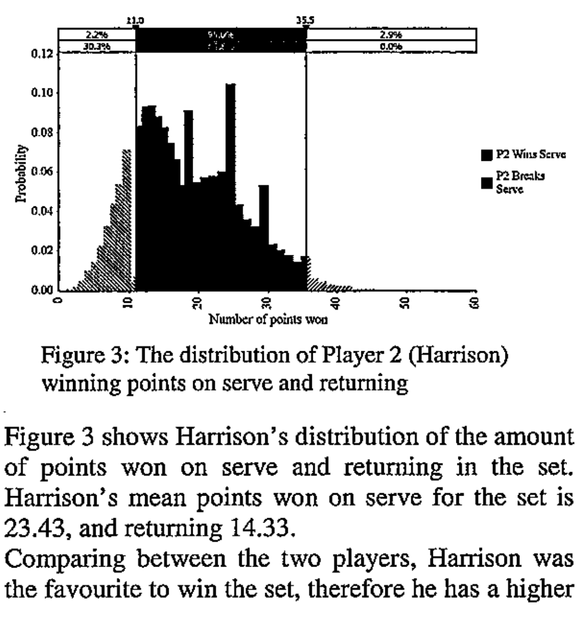


Figure 3: The distribution of Player 2 (Harrison) winning points on serve and returning

Figure 3 shows Harrison's distribution of the amount of points won on serve and returning in the set. Harrison's mean points won on serve for the set is 23.43, and returning 14.33.
Comparing between the two players, Harrison was the favourite to win the set, therefore he has a higher

5. CONCLUSION

This paper develops three methods to updating the probability of winning on serve, and comparing it to the Markov chain model whilst the game is in-play. This paper is aimed at creating a realistic element to in-play tennis to cater for a player's change in momentum, tactics, weather and injuries that will influence the probability and likelihood of winning a point on serve. Three methods and the Markov model were compared using simulated data and one actual tennis match. The results found that all three methods follow the trend of the Markov model and no optimal method could be chosen at this point. Although no chosen model was selected it appears the *Court-side* approach appears to reflect the actual phase of the game and therefore future research is required. Whilst simulating a tennis set it was found that a simulation of greater than 10,000 simulations is required in order to decrease the error rate. Future work is required to apply these findings on empirical data.

References

Barnett, T.J, Brown A. & Clarke S.R. 2006. Developing a model that reflects the outcomes of tennis matches. *Proceedings of the 8th Australasian Conference on Mathematics and Computers in Sport*,178-188.

Barnett, T. J. 2006. Mathematical modelling in hierarchical games with specific reference to tennis. *Ph.D. Thesis*, Swinburne University of Technology, Melbourne.

Klaassen, F. J. G. M. & Magnus, J. R. 2001. Are points in tennis independent and identically distributed? Evidence from a dynamic binary panel data model. *J. Am. Stat. Assoc.*, 96,500-509.

Klaassen, F. J. G. M. & Magnus, J. R 2003. Forecasting the winner of a tennis match. *European Journal of Operational Research* 148, 257-267.

HANDICAPPING IN SPORT AND RECREATION

Neville de Mestre[a, b]

[a] Bond University
[b] Corresponding author: *Neville_de_mestre@staff.bond.edu.au*

Abstract

For many sports there can be a handicap competition, usually incorporated into the main event, but not always. The Sydney-Hobart Yacht race has both a handicap section and a line-honours section, but the Stawell Gift professional foot race is based on staggered-start handicaps only, whereas most thoroughbred horse races involve weight-carrying handicaps. Three different handicap competitions will be considered here in some detail, for various reasons. Golf Australia has recently altered its handicapping system, and the resultant changes are of interest to many. Weightlifting handicap competitions are relatively unknown, and use mathematical models that have never completely satisfied the participants. Handicapping in Contract Bridge pairs events is still being developed, and offers special insight to an interesting problem.
Keywords: Golf, weightlifting, contract bridge, handicapping.

1. INTRODUCTION

There are many reasons for introducing handicaps into sporting and recreation events. It may be done to increase participation, to spread the distribution of prizes, to encourage all participants to do their best, to change the probability of winning, or simply to add variety. In this paper I will consider handicapping in golf, weightlifting and Contract Bridge, three widely-different sports and recreations.

2. HANDICAPPING IN GOLF

The aim of handicapping in golf is to make the game more inclusive and enjoyable for golfers of all abilities. The handicap system was initially devised to allow golfers to compete against each other on as equitable a basis as possible. The overriding principle in any recommended changes to the handicap system is to afford each player in a competition field a reasonable expectation of winning, or placing well, if that player plays reasonably better than his or her handicap. Australia has had three golf handicapping systems in succession in recent years. These are the old system (pre-2010), a new system introduced in 2010, and the modified new system in place since September 2011.
The Old Australian Golf Handicap system did not take into account the degree of difficulty of different courses when adjusting players' handicaps. High-handicap markers exhibit a greater standard deviation in their scores than low-handicap markers and, as a result, the old system showed a bias in favour of the low markers. The high markers won far fewer competitions than their proportional representation in the field suggested. Handicaps for high markers only extended by 0.1 of a stroke for a poor round, no matter how poor, but were reduced by a larger margin for a good round.
Consequently Golf Australia decided to change the handicapping system in 2010, and adopted in general the system used by the United States Golfing Association. This new system calculated the best ten differentials from a player's twenty most recent scores, where a differential is defined by

Differential = Gross Score – (AMCR or AWCR)

Here AMCR and AWCR are acronyms for the Australian Mens' and Womens' Course Ratings respectively. The inclusion of these in the new system was an attempt to create a fairer system for handicapping golf because of the varying difficulty of courses throughout the land. To facilitate the inclusion of the AMCR or AWCR in the calculation of each player's handicapping index, Golf Australia carried out an extensive re-rating of all courses throughout Australia using many volunteers. The ratings were based on a zero handicap for that particular course. This new Twenty10 system used the arithmetic mean of these best ten differentials , and multiplied it by a "bonus for excellence" factor

(BEF) of 0.96.This result was then rounded to the nearest integer to produce a player's handicap mark. In essence the BEF was included to help low markers. It now means that their handicaps are less likely to be reduced by the same amount than for the high markers with a series of equivalent good scores. For example, a player with a handicap of 17 who posts an average differential of 15 has that changed to 14.40 by the BEF multiplier, which produces a new handicap of 14. However a player with a handicap of 5 who posts an average differential of 3 has that changed by the BEF multiplier to 2.88, producing a new handicap of 3 only. The BEF provides an extra incentive for players to lower their handicaps.

Unfortunately it was soon felt that the new system now favoured the high markers. Golf Australia commissioned a consulting statistician, Michael Maher, to investigate this perceived problem. Using a sample of one million rounds of golf from more than 27,000 competitions involving more than 400, 000 golfers in total, he was able to compare the old and new systems in great detail and verified that the new Twenty10 system did advantage the high markers when the field sizes were large (above 50 for men and above 100 for women), but advantaged the low markers when the field size was very small (Maher, 2011a).

As a consequence of these findings, Maher (2011b) was asked to consider further options with the objective of reducing this bias. This time the data was obtained from all the home competitions of twenty-five representative clubs over a time period from December 2007 to November 2010. Although it was acknowledged that no system would be able to show no bias towards low or high markers across fields of all sizes, the following new modified Twenty8 system was adopted as the least worst result.

1. Handicaps are to be calculated using the best eight differentials of twenty most recent scores.
2. The BEF is to be 0.93.
3. A handicap cannot exceed four more than the lowest exact handicap over the previous twelve months.
4. Any differential used for handicap calculation cannot exceed 40 for men and 50 for women.

The author is grateful to Michael Maher for providing details of his two excellent and thorough reports to Golf Australia. This section of the handicapping paper is mainly a summary of Maher's analysis.

3. HANDICAPS IN WEIGHTLIFTING

Handicaps are useful in weightlifting competitions that have a small number of competitors spread over the eleven different weight classes or divisions. Weight classes are determined by body mass. As the body mass increases so does the total weight lifted. The problem for handicapping fairly is to find the relationship between body mass (B) and the expected lift (L). When the total actually lifted for the clean-and-jerk plus the snatch is represented by T, the percentage (P) of total lift to expected lift is given by

$$P = 100\, T/L$$

As an example, consider the table below which shows the total lift for each division winner in eight classes at the 2000 Sydney Olympic Games.

Name	Division (kg)	Body mass (kg)	Total lifted (kg)
Halil Mutlin	<55.99	55.62	305.0
Nikolai Peshalov	56-61.99	61.56	325.0
Galabin Boevski	62-68.99	68.78	357.5
Xugang Zhan	69-76.99	76.20	367.5
Pyrros Dimas	77-84.99	84.06	390.0
Akakios Kakisvilis	85-93.99	92.06	405.0
Hossein Tavakoli	94-104.99	104.70	425.0
Hossein Rezazadeh	>=105	147.48	472.5

Table 1

So, who was the best weightlifter based on handicap? The simplest model is to assume a linear relationship

$$L = a + c\,B$$

where a and c are constants determined from lots of weightlifting data. This turns out to be disadvantageous to both the very light and very heavy classes, because it is not based on physical theory.

Since strength is related to the cross-sectional area of the muscles, it is therefore proportional to the square of an athlete's typical linear dimension. But body mass should be proportional to the cube of an athlete's typical linear dimension yielding a power-law model

$$L = d\,B^{3/5}$$ — 2/3

where d is again determined from lots of data. Unfortunately this still has major inaccuracies for the heavier lifters. A similar model uses

$$L = f\,B^{g}$$

where f and g are constants with g very close to ⅔. By taking logarithms of both sides this essentially becomes the same as the linear model, except that the variables and constants are now the logarithms of the former ones.
Siff and Verkhoshansky (2003) used an unusual empirical relation of the form

$$L = h - j\,B^{-k}$$

with three constants h,j,k.

The three models predict three different winners on a handicap basis, showing that there is still no general agreement on the best way to determine a handicap winner in weightlifting.
The main problem seems to be that lighter lifters have a better power-to-weight ratio than others. This seems to be part of the nature and make-up of limitations of the human skeleton and muscular system, combined with the shorter leverage of smaller people.

4. CONTRACT BRIDGE HANDICAPPING

Many top bridge players are not in favour of the idea of handicapping competitions in bridge. This is probably because they usually win non-handicap events regularly at most clubs. But bridge clubs contain a lot of novice and intermediate players, and although there may be special sessions for them, it is sensible to introduce them to playing against the top players and also giving them an equal chance of winning. . Handicap events are just the vehicle for this, provided that the handicaps are fair.
Handicapping events are not set so that the lesser-experienced players always win, but so that everyone has initially an equal chance of winning.
The first step is to create a sensible expected score ES for each bridge player. This can only be based on his or her recent efforts in play and utilises recent results at their club irrespective of whom they partner, from the newest novice to a Grand Master.
Each player's results at a club are presented as a percentage between 0 and 100, with most scores sitting between 35 and 65.These scores can be collected for each player by one of a variety of special computer software packages for bridge. A reasonable expected score for each player can then be obtained by deleting the highest and lowest of the last eight scores, no matter how long these occurred over time, and averaging the remaining six. It is recommended that all newcomers to a particular club (visitors or new members) should start off initially with eight scores of 50.

Research shows that this is an excellent method and there is no need to increase or decrease the number of previous scores from eight. Some objections may be raised that a person, who has a high expected score from winning regularly in restricted events, will then find it difficult to compete against the top players in a handicap event across the whole club. That is absolutely true, and should be so. If they then play in more open matches, their expected score will come down to the correct level.
The updated expected score for each player is then averaged for each pair in a handicap competition to provide an expected handicap score EHS, where

$$EHS = (ES1 + ES2)/2$$

The expected handicap score is then subtracted from the actual score AS to produce a positive or negative number usually between -15 and +15 which shall be called the raw difference RD.

$$RD = AS - ES$$

The pair with the highest raw difference wins the handicap event.

It is true that expert players will have an expected score near 60 while novice players may be near 40. They will also have a smaller standard deviation in their scores over time. It is therefore slightly more difficult for an expert pair to obtain a positive raw difference than it is for a novice pair. To account for this, it was proposed that the raw difference be finally adjusted by a weighted excellence factor based on the probability of increasing a pair's expected score. But examination of the results for a number of handicap matches at my club in early 2012 indicates that the raw difference produces the same winners as the adjusted raw difference on all occasions bar one. This seems to show that the weighting factors may only rarely play a role in readjusting the scores to determine the handicap winners. Therefore it appears that we can forget about weighting adjustments , and just use the simple raw difference. This makes it really easy to calculate and administer. Hence for Contract Bridge handicap events:

1. Calculate each player's handicap based on the middle six of the most recent eight scores.
2. Update these at least once each week.
3. For each handicap event calculate the expected handicap score for each pair using the average (arithmetic mean) of their individual expected scores.
4. Subtract the expected handicap score from the actual score to obtain the raw difference for each pair.
5. The pair with the highest raw difference wins the handicap event

References

Maher, Michael (2011a).The Australian Golf Handicap System, 10 March 2011, 55 pages. Private communication.

Maher, Michael (2011b). Handicap Modelling Summary, May 2011, 23 pages. Private communication.

Siff, M. C. and Verkhoshansky, Y. V. (2003). Supertraining, Denver (6th edition).

EFFECTS OF CREW CHANGES ON ROWING PERFORMANCE OF OLYMPIC-CLASS FOURS AND QUADS AT INTERNATIONAL REGATTAS

Ainhoa Iñiguez Goizueta [a,d], Will G. Hopkins [b], T. Brett Smith [c]

[a] *Mathematical Institute, University of Oxford, Oxford, United Kingdom*
[b] *School of Sport and Recreation, AUT University, Auckland, New Zealand*
[c] *Department of Sport and Leisure Studies, The University of Waikato, Hamilton, New Zealand*
[d] *Corresponding author: ainiguezgoizueta@gmail.com*

Abstract

Rowing coaches often replace members of crewed boats, but the pattern of changes and effects on performance have not previously been documented. Here we present an analysis of crew changes in boats of the Olympic-class coxless fours (men's and lightweight men's) and quadruple sculls (men's and women's) in international regattas between 1999 and 2009. The data were official race times of 3-25 boats in each class in each of four regattas each year (three world cups followed by a world championship or an Olympics). There was a similar pattern of crew changes in the four boat classes, the number of changes declining gradually from 1.9 ± 1.4 (mean ± SD) in the first world cup through 0.8 ± 1.2 in world championships and 0.5 ± 0.9 in Olympics. The log of performance time was analysed with a mixed linear model that included fixed and random effects to adjust for environmental and other factors affecting performance while providing percent estimates of mean change in performance time and any individual differences in the change when there was a change in the crew. Inferences were based on uncertainty in effects in relation to a smallest important change in performance of ~0.3%. A change of 1-2 crew members generally had little mean effect on performance, with at most only small individual differences representing successful or unsuccessful changes. There was a small mean decline in performance and large individual responses with one change in the coxless four, possibly arising from errors in the official database. All boat classes showed small mean declines in performance (0.3-0.6%) and evidence of substantial individual differences with a completely new crew. We conclude that rowing coaches need have little concern about making several changes to an existing crew in the Olympic fours and quads, but new crews will generally underperform.

Keywords: Elite athletes, competition, mixed modelling, Olympics

1. INTRODUCTION

Rowing coaches often replace members of crewed boats between competitions for reasons such as illness, injury, poor performance, or retirement of one or more crew members. In our experience of the sport, coaches make fewer changes closer to important competitions, presumably because of concern that a change could result in impaired performance for at least the next competition. However, the pattern of changes in crewed boats during a competitive season and the effects of changes on performance have not been documented previously. The aim of this study was to analyse the changes and their effects on performance using a database compiled for a recent study of factors affecting rowing performance in international regattas (Smith & Hopkins, 2011).

2. METHODS

The database consists of official performance times of all boats in all international regattas between 1999 and 2009. There were four regattas each year: three world cups followed by either a world championship or an Olympics. We limited the analysis to Olympic-class coxless fours (men's and lightweight men's) and quadruple sculls (men's and women's). Names of all crew members were included in the database, so it was possible to add a variable denoting the number of changes in crew members in every boat since the previous regatta. We identified and tracked boats through the regattas by nation. For the rare occasions when nations entered two boats in the same regatta, we scrutinized contiguous regattas and chose the nation's main boat on a case-by-case basis; the other boat was then deleted from the analysis.

Data were compiled and analysed with the Statistical Analysis System (Version 9.2, SAS Institute, Cary, NC). Percent effects of crew changes on performance time were analyzed via log transformation using a mixed linear model (Proc Mixed) with the number of changes in the boat (five levels: 0 through 4) as a fixed effect; other fixed and random effects were included to adjust for the stage of competition (five levels: world cups 1 through 3, world championship, Olympics), the class of the final (up to six levels, A through F), the mean speed of each boat within and between years, environmental effects within and between regattas, and the venue effect (Smith & Hopkins, 2011). Separate residual variances for each number of changes in the boat provided estimates of within-boat regatta-to-regatta variation that included individual responses to the effect of crew changes. Four observations were identified as outliers (standardized residuals >4.5) in a first run through the analysis and were eliminated before re-analysis.

Inferences were based on uncertainty in magnitude of effects in relation to a smallest important change in performance of ~0.3% (equivalent to one extra win or loss for a best boat every 10 regattas) (Hopkins, Marshall, Batterham, & Hanin, 2009; Smith & Hopkins, 2011). Uncertainty in estimates is shown as 90% confidence intervals, and magnitudes of effects were interpreted probabilistically (Hopkins et al., 2009).

3. RESULTS

The number of nations competing in each of the four boat classes over the 11-year period ranged from 20 (women's quad sculls) to 33-35 (men's events), and the number of boats in each class in each regatta ranged from 3 to 25. The pattern in the number of changes in the crew each year was similar in the four boat classes: most changes occurred in the first world cup each year, and the number of changes declined gradually during the year, ending with fewest changes in the world championships and Olympics. The changes averaged over the four boat classes and 11 years are shown in Figure 1.

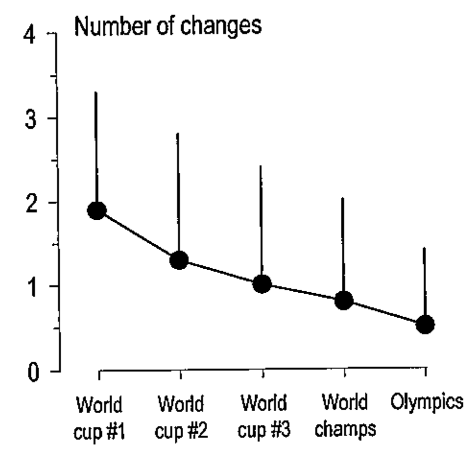


Figure 1: Number of changes occurring in each regatta averaged over all four boat classes in all years. The changes in the first world cup represent changes from the composition of the crews in each boat's last regatta in the previous season. Bars are standard deviations.

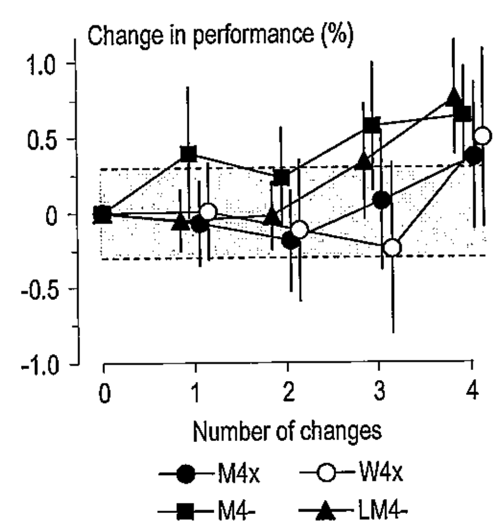


Figure 2: Change in performance associated with changes of 1-4 crew members, relative to performance when there were no crew changes. Shaded area represents trivial changes in performance (±0.3%). Bars are 90% confidence intervals. M4x, men's quad sculls; W4x, women's quad sculls; M4-, men's coxless four; LM4-, lightweight men's coxless four.

Figure 2 shows the changes in performance as the number of changes in the crew increased. The men's coxless four was the only boat class showing the possibility of a substantial decline in performance (increase in performance time) with one change, the other classes showing changes that were likely or

very likely trivial. There was no substantial mean decline in performance for two changes, but there was some possibility of substantial decline for three changes, and all boat classes showed substantial and clear mean declines for four changes.
The residual variations in performance for each number of changes (representing a boat's race-to-race variation for the given number of changes) are shown in Figure 3. Although there is considerable uncertainty in the estimates, the pattern is consistent with an overall increase in variability for three changes and especially for a complete change of crew, but little increase with up to two changes. The exception is the men's coxless four, which showed the greatest variability with a single change of crew. Examination of plots of residuals vs predicted for this boat class did not reveal any obvious pattern of non-uniformity or marginal outliers.

Environmental and other factors included in the model to adjust for their effects on performance had means and/or standard deviations similar to those in Smith and Hopkins (2011) (data not shown).

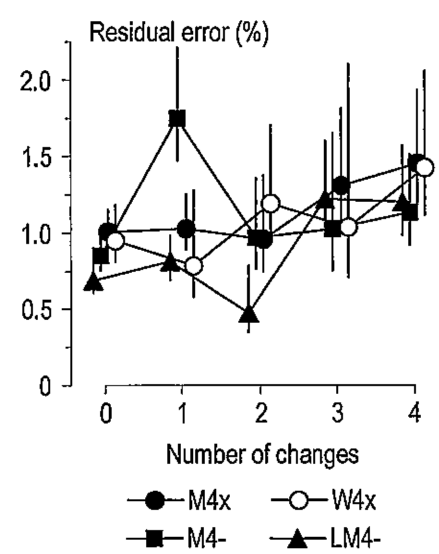


Figure 3: Residual error (representing a boat's race-to-race variation) for each number of changes in the crew. Bars are 90% confidence intervals. For abbreviations see Figure 2.

4. DISCUSSION

We observed the expected reduction in the number of changes in the boats as the rowing season progressed, with the least number of changes (approximately one change per two boats) in the Olympics. Coaches are obviously concerned about making changes before important competitions, or are more willing to experiment with crews earlier in the season.

For three of the boat classes (M4x, W4x, LM4-), a change of one or two crew members had no substantial mean effect on performance. Some of these changes would have resulted in replacements with rowers who would be better or worse as individuals, but the analysis of residual variability of the new crew provided no evidence of substantial improvements and impairments of performance. Evidently any substantial difference in ability between one or two replacements and the existing crew was diluted by the ability of the existing crew. The men's coxless four (M4-) did not fit this pattern, and the only explanation that seems reasonable to us is a subtle error or errors in the database that did not produce outliers, such as mislabelling of rowers.

With a change of three rowers, and especially with a complete change of crew, all boat classes showed a possibility of impaired performance overall and substantial individual differences in the impairment. It follows that most major crew changes will produce impairments in performance at the next regatta.

5. CONCLUSIONS

A change of 1-2 crew members generally had little mean effect on performance with at most only small individual differences representing successful or unsuccessful changes. All boat classes showed small mean declines in performance and evidence of substantial individual differences with a completely new crew. We conclude that rowing coaches need have little concern about making several changes in an existing crew in the Olympic fours and quads, but new crews will generally underperform.

References

Hopkins, W. G., Marshall, S. W., Batterham, A. M., & Hanin, J. (2009). Progressive statistics for studies in sports medicine and exercise science. Medicine and Science in Sports and Exercise, 41(1), 3-12.

Smith, T. B., & Hopkins, W. G. (2011). Variability and predictability of finals times of elite rowers. Medicine and Science in Sports and Exercise, 43, 2155-2160.

A GENERAL MEASURE FOR THE RELATIVE EFFICIENCY OF ANY TWO SCORING SYSTEMS

Pollard, Graham [a, c], Pollard, Geoff [b]
[a] *Faculty of Information Sciences and Engineering, University of Canberra, Australia*
[b] *Faculty of Life and Social Sciences, Swinburne University of Technology, Melbourne, Australia*
[c] *Corresponding author: graham@foulsham.com.au*

Abstract

Miles (1984) developed a very elegant theory for the relative efficiency of different scoring systems at correctly identifying the better player, assuming points were independent. This earlier work was limited to those situations in which the underlying probability structures of the game being modelled had certain restrictive characteristics. Using those underlying characteristics it was possible to use interpolation methods to derive efficiency measures in a restricted number of practical situations.

The major objective of this research was to investigate whether Miles' work on the efficiency of scoring systems could be extended to more general situations. Games that do no possess the restrictive probability structures noted above have been considered, and it has been shown that an extrapolation method for deriving efficiency measures can be developed and applied. In doing so the efficiency of nested scoring systems has been studied.

It turns out that this extrapolation method can be used in any scoring system situation, even where the outcome is win/draw/loss rather than win/loss. It produces exactly the same efficiency formula as that produced by the interpolated method. Thus, the method for measuring efficiency has been extended to a wider range of practical scoring systems situations.

Keywords: interpolated efficiency, extrapolated efficiency, constant probability ratio property, (P, μ, n) equations, efficiency of nested scoring systems, relative efficiency of statistical sequential probability ratio tests, Win-by-n scoring systems

1. INTRODUCTION

Miles (1984) developed a very elegant theory for the relative efficiency of different scoring systems, assuming that points were independent. He considered 'win-by-n' (Wn) scoring systems in which the winner was the first player to win n more points than his opponent. Assuming points are independent and player A has a constant probability p of winning every point (unipoints), it can be shown that the probability P that player A wins Wn, the expected number of points played μ, and n satisfy the '(P, μ, n) equation'

$$(P-Q)/\mu=(p-q)/n$$

and that the ratio P/Q is given by

$$(P/Q)=(p/q)^n$$

where Q = 1 − P and q = 1 - p. These equations follow from the fact that W_n has the constant probability ratio (cpr) property (Pollard, 1992). That is, the ratio of the probability that player A wins in n + 2m points divided by the probability that he loses in n + 2m points (m = 0, 1, 2, ...) is constant, and is equal to P/Q. Noting the optimal nature of this Wn system (Wald and Wolfowitz, 1948), and using Wn (points) as the family of scoring systems with unit efficiency, Miles (1984) showed that the efficiency ρ of a general 'unipoints' scoring system SS with key characteristics P and μ is given by

$$\rho = \frac{(P-Q)\ln(P/Q)}{\mu(p-q)\ln(p/q)} \quad (1)$$

This efficiency measure, as described by Miles (p. 97), is defined as the expected duration of the 'interpolated' Wn system with the same P-value as SS (as derived from the above '(P, μ, n) equation') divided by the expected duration of SS, namely μ. Note that the value of n for this 'interpolated' Wn system is given by

$$n = \ln(P/Q)/\ln(p/q),$$

resulting from the cpr property of Wn.

It follows, by ignoring the factors involving p (and q) in (1) above, that the expression

$$((P-Q)/\mu)\ln(P/Q)$$

is the measure for the relative efficiency *of a unipoints scoring system given underlying independent points.*

Miles (1984) also considered scoring systems relevant to tennis (and other sports such as volleyball), which he called 'bipoints' scoring systems. He assumed that the probability player A (B) wins a point on service is p_a (p_b), and that points are independent. Noting the work of Wald (1947) and using Wn (point-pairs) as the standard family of scoring systems with unit efficiency, he showed that the efficiency of a general bipoints scoring system with key characteristics P and μ is given by

$$\rho = \frac{2(P-Q)\ln(P/Q)}{\mu(p_a - p_b)\ln(p_a q_b / p_b q_a)} \quad (2)$$

where $q_a = 1 - p_a$ and $q_b = 1 - p_b$. As in the unipoints Wn case, the Wn(point-pairs) family of scoring systems possesses a '(P, μ, n) equation' and a cpr property, leading directly to equation (2) by using the same 'interpolation' method as that used above for unipoints.

Thus, ignoring the constant and the factors involving p_a and p_b, the measure for relative efficiency is *also* given by

$$((P-Q)/\mu)\ln(P/Q)$$

for this bipoints system *with underlying independent points with constant p-values p_a and p_b*. That is, the efficiency of the bipoints scoring system 1 relative to the bipoints scoring system 2 is given by

$$\frac{((P_1-Q_1)/\mu_1)\ln(P_1/Q_1)}{((P_2-Q_2)/\mu_2)\ln(P_2/Q_2)} \quad (3)$$

using an obvious notation.

Pollard and Pollard (2008) used the above interpolation method to show that (3) is also the measure for relative efficiency for the independent quad-points case (e.g. tennis doubles with parameters p_{a1}, p_{a2}, p_{b1} and p_{b2}). Further, they showed (Pollard and Pollard, 2010), again using the interpolation approach, that (3) is the relative efficiency expression for scoring systems where unipoints or bipoints become one-step dependent probabilities.

It was possible to derive the relative efficiency for each of these four situations (unipoints, bipoints, quadpoints (e.g. tennis doubles), 1-step dependent unipoints and bipoints) because in each case the underlying point probability structure of the situation being modelled lead to both a '(P, μ, n) equation' and cpr property for the relevant underlying Wn system. This however is not always the case, and the interpolation approach is not possible when it is not. For example, supposing player A has a probability p of winning a point when the players are equal, p+ when he is ahead, and p- when he is behind, it can easily be seen that the Wn system of scoring systems does not have the cpr property when n > 2.

In this paper we consider an alternative approach to relative efficiency. This alternative approach does not depend on there being a '(P, μ, n) equation' (and a cpr within an underlying Wn system) that is necessary for the interpolation method.

2. METHODS

Miles (1984) noted that efficiency under nesting was 'roughly multiplicative' (p. 107). An aspect of this approximation is that the expected number of points in a set of tennis is only approximately equal to the expected number of points in a game (different for each player) multiplied by the expected number of games in a set (Pollard, 1983). Further, the expected number of service games for one player is typically different to the expected number for the other. *We now consider several examples in which efficiency under nesting is exactly multiplicative.*

Example 1

Suppose player A has a constant probability 0.6 of winning a point and that points are independent. Consider the nested system B3(B3), where the outer nest represents a 'set', and the inner nest represents a 'game'. Note that the nested system can be won or lost by player A in as few as 4 points, or as many as 9 points. First principles can be used to show that this nested system has a mean of 6.09135616 points, a probability player A wins of 0.715516416, giving an efficiency of 0.8048199542 (by using (1)). The inner nest has a mean of 2.48 points, and a probability that player A wins of 0.648, giving an efficiency of 0.8981959879 (using (1)), whilst the outer nest with a p-value of 0.648 has a mean of 2.456192 games, and a probability player A wins of 0.715516416, giving an efficiency of 0.8960404689. It can be seen that the product of these efficiencies for the inner and outer nests is exactly equal to the efficiency of the total system, as calculated above.
The efficiency of the inner nest can be expressed as

$$\rho_i = \frac{(P_i - Q_i)\ln(P_i / Q_i)}{\mu_i(p-q)\ln(p/q)}$$

and the efficiency of the outer nest as

$$\rho_o = \frac{(P_o - Q_o)\ln(P_o / Q_o)}{\mu_o(P_i - Q_i)\ln(P_i / Q_i)}$$

using obvious notations, whilst the efficiency of the total nested scoring system is

$$\rho_n = \frac{(P_n - Q_n)\ln(P_n / Q_n)}{\mu_n(p-q)\ln(p/q)}$$

Noting that Pn is the same as Po (and Qn is the same as Qo), it follows that $\rho_n = \rho_i * \rho_o$ if $\mu_n = \mu_i * \mu_o$.

Thus, *the efficiency of a nested system is exactly multiplicative when the expected duration of the nested system is exactly equal to the product of the expected durations of the nests. It is clear that this also applies to triple nesting, etc.*
Some other unipoints examples where efficiency under nesting is exactly multiplicative are other B2n-1(B2m-1) systems such as B3(B5), Wn(B2n-1) systems such as W2(B3), and Wn(Wm) systems such as W2(W3).
It is clear that 'exact multiplicative efficiency' also applies in bipoints, quadpoints, etcetera, when the 'means are multiplicative', as the form of the efficiency expressions remains the same.

Example 2

We consider the nested system Wn(SS) where SS is a scoring system with probability player A wins equal to p, mean duration equal to μ points, mean duration conditional on player A winning equal to μ_W points and mean duration conditional on player A losing equal to μ_L. Suppose D_z is the expected number of points remaining in the nested system when z is the score of the outer nest (z = -n, -n + 1, -n + 2,..., n - 1, n) and an inner nest is about to begin. It is clear that $D_n = 0$ and $D_{-n} = 0$ are boundary conditions. We have the recurrence relations

$$D_z = p(D_{z+1} + \mu_W) + q(D_{z-1} + \mu_L) \text{ i.e.}$$

$$D_z = pD_{z+1} + qD_{z-1} + \mu$$

where q = 1 - p. It follows using the methods described in Feller (1950, p. 317) that

$$D_z = \frac{\mu(z-n)}{q-p} + \frac{2n\mu((q/p)^n - (q/p)^z)}{(q-p)((q/p)^n - (q/p)^{-n})}$$

Putting z = 0, it follows that the expected duration of the nested system is given by

$$D_0 = \frac{P-Q}{p-q} n\mu$$

where

$$P = p^n / (p^n + q^n)$$

and

$$Q = 1 - P.$$

Thus, *the mean of the nested system Wn(SS) is equal to the mean of the inner nest SS multiplied by the mean of outer nest (i.e. the mean of Wn at p). Also,* *the efficiency of Wn(SS) is equal to the efficiency of SS, since the efficiency of the Wn system is unity.*

Example 3

As a special case of Example 2, W4(B3) with point probability 0.6 is considered. In this case p = 0.648, q = 0.352, μ = 2.48 and D_0 = 28.14489049 using the above equation. This value of D_0 was verified using standard recurrence methods with inner nest conditional means of μ_W = 2.444444 points and μ_L = 2.545454 points. (In the process of considering the state of the outer nest after every second inner nest was completed, it is noted (as a bi-product) that D_2 was equal to 15.53354219 points and D_{-2} was equal to 30.85308066 points, and these values agree with the above equation for D_z.)
Example 3 is a unipoints one. A bipoints example in which the mean of the nested system is exactly equal to the product of the mean of the inner nest and the mean of the outer nest, would be a 'set' of tennis defined as W_2(TB) where TB is the usual tiebreak game.
The above expression for D_0 is used in the following section to extend our definition of efficiency to the situation in which a '(P, μ, n) equation' does not exist for the underlying probabilistic structure under consideration.

Extended definition of relative efficiency

Suppose the two scoring systems SS1 and SS2 have identical underlying probabilistic structures, and that SSi has an expected duration of μ_i points and a probability that player A wins of p_i (i = 1, 2).
Consider the nested scoring systems Wn1(SS1) and Wn2(SS2). The probability player A wins Wni(SSi), Pi (i = 1, 2), can be evaluated using the relationship

$$P_i / Q_i = (p_i / q_i)^{n_i}$$

where $P_i + Q_i = 1$ and $p_i + q_i = 1$,
and the expected duration of Wni(SSi) is equal to

$$((P_i - Q_i)/(p_i - q_i))n_i\mu_i,$$

using the above equation for D_0.

Now suppose n1 and n2 are two (possibly very large) values such that player A has the same probability of winning under either nested system. That is,

$P_1 = P_2$ and hence $P_1 / Q_1 = P_2 / Q_2$, and
$P_1 - Q_1 = P_2 - Q_2$.
It follows that

$$n_1 / n_2 = (\ln(p_2 / q_2)) / \ln(p_1 / q_1).$$

Using the underlying concept of efficiency and noting that $P_1 = P_2$ for the two nested systems, the efficiency of the system Wn1(SS1) relative to the system Wn2(SS2) is given by the mean of Wn2(SS2) divided by the mean of Wn1(SS1). That is, it is given by the expression

$$\frac{((p_1 - q_1)/\mu_1)\ln(p_1 / q_1)}{((p_2 - q_2)/\mu_2)\ln(p_2 / q_2)}. \qquad (4)$$

Since in general the efficiency of Wn(SS) is equal to that of SS, it follows that *the efficiency of the system SS1 relative to SS2 is given by expression (4).* Thus, the expression for the relative efficiency for this case (where a '(p, μ, n) equation' does not necessarily exist) is identical to (3) (for the case when the '(p, μ, n) equation' and the cpr do exist). That is, our measure of relative efficiency is no longer limited to the situation where the underlying probability point structure necessarily allows a Wn system with the cpr property and a '(p, μ, n) equation' to be established. Thus, *the relative efficiency of two systems can now be measured in a much broader range of situations than earlier (and using the same expression).* In comparison to the 'interpolation approach' to relative efficiency, *the above approach might be called the 'extrapolation' approach to relative efficiency.*

Example 4

In this example the interpolation and extrapolation methods are shown to give identical results.
Suppose player A's probability of winning a point is 0.6 and points are independent. For the scoring system SS1 = best of 3 points = B3, we have p_1 = 0.648 and μ_1 = 2.48 points and for the scoring system SS2 = best of 5 points = B5, we have p_2 = 0.68256 and μ_2 = 4.0656 points, and the above equation becomes

$$n_1 / n_2 = 1.254485322.$$

Thus, as an example of two nested scoring systems, player A has the same probability of winning W1,000,000,000(B5) as he does of winning W1,254,485,322 (B3). [These particular nested scoring systems with very very large expected durations have been chosen so that we have (more than) ample accuracy to satisfactorily demonstrate the extrapolation approach. Note that in practice we don't need n1 and n2 to be anywhere near as large in order to achieve sufficient accuracy.] Using the above equation for D_0, W1,254,485,322 (B3) has expected duration

$$((P_1-Q_1)/(0.648-0.352))*3.111123599*10^9$$

points, and W1,000,000,000(B5) has expected duration

$$((P_2-Q_2)/(0.68256-0.31744))*4.0656*10^9$$

points where $P_1-Q_1=P_2-Q_2$ are each of course extremely close to unity.

Thus, the efficiency of W1,000,000,000(B5) relative to W1,254,485,322(B3), and hence the efficiency of B5 relative to B3, when the point probability is 0.6, being the ratio of the above two expected durations, is equal to 0.943922915.

It can be seen that, using the interpolation approach, the relative efficiency expression

$$((P-Q)/\mu)\ln(P/Q)$$

is equal to 0.07283742667 for B3, and is equal to 0.06875291605 for B5, so that the efficiency of B5 relative to B3, given by the ratio of these two numbers, is equal to 0.943922915 when p = 0.6, in agreement with the above calculations. Thus, the extrapolation method and the interpolation method give identical results.

Example 5

In this example the interpolation approach to efficiency is not available, but we can use the 'extrapolation method'.

Suppose player A has a probability 0.7 of winning a point when ahead, a probability 0.6 of winning a point when equal, and a probability 0.5 of winning a point when behind. Given this underlying probability structure, it can be seen that the associated family of scoring systems W3, W4, ...do not have a overarching general '(P, μ, n) equation' nor the cpr property. For example, for W3, the probability player A wins in 3 points divided by the probability he loses in 3 points is equal to 2.94, whereas the probability player A wins in 5 points divided by the probability he loses in 5 points is equal to 2.75333, indicating that a cpr does not exist for this underlying probability structure.

Given the above underlying probability structure when player A is ahead, equal and behind, the probability player A wins a best of 3 points game is equal to 0.648 and the expected duration of the game is 2.38 points. Also, the probability he wins a best of 5 points game is equal to 0.68112 and the expected duration of the game is 3.8382 points. Thus the expression

$$((P-Q)/\mu)\ln(P/Q)$$

is equal to 0.07589782275 for B3, and it is equal to 0.07162537163 for B5, and so the efficiency of B5 relative to B3 is equal to 0.9437078566. Note that this relative efficiency is a slightly different value to that in the previous example, not surprisingly as the underlying probability values and structures are slightly different.

Further extension of relative efficiency to general Win-Draw-loss scoring systems

We now consider two scoring systems SS1 and SS2 which have expected durations μ_1 and μ_2 points and which can result in a win, a draw, or a loss to player A with probabilities p_i, d_i, q_i respectively ($p_i + d_i + q_i = 1$, $i = 1, 2$). Each of these two scoring systems can be converted to one which must result in a win or a loss to player A by repeatedly using the system until a draw does not occur. [Note that this is similar to the structure of W1(point-pairs) in bipoints.] Such systems can be represented by W1(SSi). This very natural conversion from two win/draw/loss systems to two win/loss systems produces scoring systems with expected durations equal to $\mu_i/(1-d_i)$ ($i=1, 2$). The probability player A wins under this converted system W1(SSi) is clearly equal to $p_i/(1-d_i)$ and the probability he loses is equal to $q_i/(1-d_i)$, and it follows from (4) above that the efficiency of W1(SS1) relative to W1(SS2) is equal to

$$\frac{((p_1 - q_1)/\mu_1)\ln(p_1/q_1)}{((p_2 - q_2)/\mu_2)\ln(p_2/q_2)} \quad (5)$$

as the various (1-d_i) elements above cancel out in expression (4). Note that expression (5) applies to the situation in which draws are possible, whilst expression (4) is for the case in which draws are not possible. Also, note that the draw probabilities d_i are absent from expression (5).

Interestingly, this result is related to earlier work on the asymptotic efficiency of some (statistical) sequential probability ratio tests, SPRTs (or Wn systems with n large) which can be decomposed into small independent components ,called 'modules' (Pollard, 1990). These modules were equivalent to steps in a random walk, Z_i, which were independent variables on the integers ..., -2, -1, 0, 1, 2, ... Using the approach of Cox and Miller (1965, pp 46-58), the moment generating function of Z_i is defined by

$$f(\theta) = \sum_{j=-\infty}^{\infty} e^{-j\theta} P(Z_i = j),$$

and $\theta = 0$ is clearly one root of the equation

$$f(\theta) = 1.$$

If $E(Z_i) \neq 0$, there is a unique real second root $\theta_0 \neq 0$ which has the same sign as $E(Z_i)$. (If $E(Z_i) = 0$, $\theta = 0$ is a double root). Pollard showed that the *asymptotic* efficiency of SPRT1 (with module 1) relative to SPRT2 (with module 2) is equal to

$$\frac{\theta_{0,1} E(Z_1)/E(D_1)}{\theta_{0,2} E(Z_2)/E(D_2)} \quad (6)$$

where D_i is the expected duration of module i. Noting that for the scoring systems SS1 and SS2 under consideration in this section

$$E(D_i) = \mu_i,$$

$$E(Z_i) = p_i - q_i \text{ and}$$

$$f(\theta_i) = p_i e^{-\theta_i} + d_i + q_i e^{\theta_i}, \text{ and so we have}$$

$$\theta_{0,i} = \ln(p_i/q_i).$$

Thus, the asymptotic relative efficiency of these two quite general scoring systems (using the module approach and given by (6)) is identical to the non-asymptotic relative efficiency given by (5).

An application of Win-Draw-Loss structures to tennis

'Game-pairs' with the win/draw/loss structure form an important 'building block' within tennis scoring systems. *In this next example we demonstrate how the efficiency of two alternative components within a scoring system can be directly compared without the need to assess the two full alternative systems in their entirety.*

Example 6

Here we consider the efficiency of a 'game-pair' using advantage tennis games relative to a 'game-pair' using '50-40' games (Pollard and Noble, 2004). In the '50-40' game, in order to win the game the server has to reach 50 (one more point than 40) before the receiver reaches 40. The receiver only needs to reach 40 in order to win the game.

Suppose player A has a point probability on service of 0.7, and player B has a point probability on service of 0.6. Using advantage games player A has a probability of 0.900788966 of winning a game on service, and the game has an expected duration of 5.831489655 points, whilst player B has a probability of 0.735729231 of winning a game on service, and the game has an expected duration of 6.484184615 points. For '50-40' games these values are respectively 0.74431, 4.9579 points, 0.54432 and 4.9728 points.

Thus, for the advantage game-pair (p, d, q) is equal to (0.2380521928, 0.6889553495, 0.07299245775) and μ = 12.31567427. For the '50-40' game-pair (p, d, q) is equal to (0.3391671808, 0.5216556384, 0.1391771808) and μ =9.9307 points. Using (5) it follows that the efficiency of W1('50-40' game-pairs) relative to W1(advantage game-pairs) is equal to 1.132224066 when (pa, pb) = (0.7, 0.6). That is, '50-40' games are about 13% more efficient when (pa, pb) = (0.7, 0.6).

Not only is the '50-40' game-pair more efficient as a construct than is the advantage game-pair at (0.7, 0.6), but it has a smaller variance of duration. This is an attractive property of the '50-40' game since it is often the case that the more efficient of two systems has the disadvantage of having a larger variance of duration.

Thus, the '50-40' game is particularly relevant to men's doubles, as the point p-values for men's doubles average 0.65 or more.

Example 7

We finish this section with an example of using the methods in this paper to explain why the 'play-the-loser' (PL) service exchange mechanism is more efficient than 'play-the-winner' (PW) when service is an advantage, as in tennis. Using gpp to represent 'general point-pairs' (as in Pollard, 1992), consider the scoring system Wn(PLgpp). Here the match starts with an ab point-pair, a point-pair lost by player A is followed by the point-pair aa, a point-pair won by player A is followed by the point-pair bb, and a drawn point-pair is followed by the point-pair ab, and the match is won by the first player to be 2n points ahead. For this system we have, using an obvious notation,

$$\frac{P_{PL}}{Q_{PL}} = \frac{p_a q_b^{2n-1}}{p_b q_a^{2n-1}}$$

and

$$\frac{\mu_{PL}}{P_{PL} - Q_{PL}} = \frac{2(1+(n-1)(p_a+p_b))}{p_a - p_b}$$, where p_a, p_b, q_a and q_b have been defined earlier.

For the associated system Wm(PWgpp), we have

$$\frac{\mu_{PW}}{P_{PW} - Q_{PW}} = \frac{2(1+(m-1)(q_a+q_b))}{p_a - p_b}$$, and

$$\frac{P_{PW}}{Q_{PW}} = \frac{p_a^{2m-1} q_b}{p_b^{2m-1} q_a}.$$

Now suppose we consider two such systems with

$$\frac{P_{PL}}{Q_{PL}} = \frac{P_{PW}}{Q_{PW}}.$$

It follows that

$(q_b / q_a)^{n-1} = (p_a / p_b)^{m-1}$, and hence, using the expansion for ln((1+x)/(1-x)), we have

$$\left(\frac{(n-1)p}{(m-1)q}\right)\left(\frac{1+(\delta/q)^2/3}{1+(\delta/p)^2/3}\right) = 1$$

where $p = (p_a + p_b)/2$, $q = 1 - p$ and $\delta = p_a - p$, and powers of δ^4 and higher are omitted. The second expression in brackets is greater than 1 in the tennis context ($p > 0.5$), and so the first expression must be less than 1. It follows that μ_{PL} is less than μ_{PW}, and so the PL system is the more efficient, as their P-values are equal. Correspondingly, the PW system is the more efficient when $p < 0.5$.

3. RESULTS

Suppose two players or two teams are playing a sport and there are two scoring systems (each with a win/loss outcome) under consideration for use, namely SS1 and SS2. Suppose SSi has an expected duration of μ_i points, the better player or team has a probability of p_i of winning under SSi ($i = 1, 2$), and a probability of q_i of losing (here $q_i = 1 - p_i$). It has been shown that *under these very general assumptions*, the efficiency of SS1 relative to SS2 is given by the ratio

$$\frac{((p_1 - q_1)/\mu_1)\ln(p_1/q_1)}{((p_2 - q_2)/\mu_2)\ln(p_2/q_2)}.$$

If the outcome under each scoring system SS1 and SS2 is instead win/draw/loss with probabilities $p_i/d_i/q_i$, then the efficiency of repeatedly playing SS1 until one player or team wins [namely W1(SS1)] relative to repeatedly playing SS2 until one player or team wins [namely W1(SS2)] is given by the same ratio.

It is clear that if SS1 is more efficient than SS2, and SS2 is more efficient than SS3, it follows that SS1 must be more efficient than SS3. Thus, the most efficient of a set of scoring systems can be identified.

4. CONCLUSIONS

Earlier work on the efficiency of scoring systems has been limited to those situations in which the underlying probability structures for the game being modelled had certain restrictive characteristics. Using those underlying characteristics it was possible to use interpolation methods to derive efficiency measures.

In this paper games that do no possess such restrictive probability structures have been considered, and it has been shown that extrapolation methods for deriving a relative efficiency measure can be developed and applied.

It turns out that that this extrapolation method can be used in many scoring system situations, and it produces exactly the same efficiency formula as that produced by the interpolated method. Thus, the method for measuring efficiency has been extended to a wider range of probabilistic situations.

The efficiency of nested scoring systems, whilst roughly multiplicative for the present tennis scoring system(s), has been shown to be exactly multiplicative for many situations.

References

Cox, D. R. & Miller, H. D. (1965) The theory of Stochastic Processes. London: Chapman and Hall.

Feller, W. (1957) An introduction to Probability Theory and its applications. John Wiley & Sons.

Miles, R. (1984). Symmetric sequential analysis: the efficiencies of sports scoring systems (with particular reference to those of tennis). *Journal of the Royal Statistical Society B*. 46(1), 93-108.

Pollard, G. H. (1983). An analysis of classical and tie-breaker tennis. *Australian Journal Statistics, 25(3)*, 496-505.

Pollard, G. H. (1990). A method for determining the asymptotic efficiency of some sequential probability ratio tests, *The Australian Journal of Statistics*, 32, 191-204.

Pollard, G. H. (1992). The optimal test for selecting the greater of two binomial probabilities, *The Australian Journal of Statistics*, 34, 273-284.

Pollard, G. H. and Noble, K. J. (2004). The benefits of a new game scoring system in tennis: the 50-40 game 50-40 Game In *Proceedings of the Seventh Australasian Conference on Mathematics and Computers in Sport*, edited by R. H. Morton & S. Ganesalingam, Massey University, Massey, New Zealand, 262-265.

Pollard, G. N. and Pollard, G. H. (2008). The efficiency of doubles scoring systems. *Proceedings of the Ninth Australasian Conference on Mathematics and Computers in Sport*, edited by J. Hammond, Mathsport (ANZIAM), 45-51.

Pollard, G. H. and Pollard, G. N. (2010). The efficiency of scoring systems with outcome dependent points. *Proceedings of the Tenth Australasian Conference on Mathematics and Computers in Sport*, edited by A. Bedford and M. Ovens, Mathsport (ANZIAM), 35-42.

Wald, A. and Wolfowitz J. (1948). Optimal Character of the sequential probability ratio test. Ann. Math. Statist. 19, 326-339.

Wald, A. (1947) Sequential Analysis. New York: Wiley.

MATCH FIXING AT THE LONDON 2012 OLYMPIC GAMES

Yvette Mojica Perez, Anthony Bedford, & Adrian J Schembri

[a] *School of Mathematical and Geospatial Sciences, RMIT University*

[b] *Corresponding author: anthony.bedford@rmit.edu.au*

Abstract

Match fixing has occurred in many sports from the amateur leagues all the way up to the professional teams. The aim of this paper was to examine any potential match fixing incidents at the London 2012 Olympics and assess whether match scheduling could minimise match fixing in the future. The results and schedules for the badminton, football and basketball events were assessed. Match fixing was evident in each of these sports, with the badminton event being the most publicised due to the disqualification of four women's doubles teams. Inspection of the match schedules used in each of the events revealed that match scheduling could be used to reduce match fixing. In badminton a match schedule which included the top ranked teams playing in the first round could minimise match fixing. Meanwhile for the football events, the top two ranked teams playing against each other in the final group match could provide fewer opportunities for match fixing.

Keywords: Match fixing, Olympic games,

1. INTRODUCTION

Match fixing is prominent in the history of sports such as soccer, basketball, tennis and cricket (McLaren, 2008; Preston & Szymanski, 2003) and has returned with the latest and highly publicised incidents occurring at the London 2012 Olympics. Match fixing occurs when an individual or team manipulates the outcome of a particular match. The definition of match fixing as agreed by the Australian Government (p.2, 2011) National Policy on Match-Fixing in Sport is as follows: "Match-fixing involves the manipulation of an outcome or contingency by competitors, teams, sports agents, support staff, referees and officials and venue staff. Such conduct includes:

a. the deliberate fixing of the result of a contest, or of an occurrence within the contest, or of a points spread;
b. deliberate underperformance;
c. withdrawal (tanking);
d. an official's deliberate misapplication of the rules of the contest;
e. interference with the play or playing surfaces by venue staff; and
f. abuse of insider information to support a bet placed by any of the above or placed by a gambler who has recruited such people to manipulate an outcome or contingency."

There are many reasons why individuals or teams partake in match fixing. Match fixing may be motivated by financial gains and obtaining a better draw in the knockout rounds (either to benefit one team/player or both teams/players involved) (Preston & Szymanski, 2003). Tournament design can influence individuals' or teams' desire to match fix a certain result if it benefits the individual or team (Preston & Szymanski, 2003). One method that can be used to minimise incidents of match fixing is organising the scheduling of matches to be played. Types of match scheduling coupled with the standard deviation of each team's ranking in a group have been found to be associated with the potential of match fixing (Schembri, Golkhandan,

& Bedford, 2010). Schembri et. al. found that incidences of match fixing could be minimised if the two highest ranked teams in a group played against each other in the last group match. They also found that groups with a larger standard deviation in team Elo ratings were potentially more likely to result in match fixing. On the other hand, groups consisting of teams that are very closely ranked were less likely to result in match fixing.

A study conducted by Page and Page (2009) investigated the group stages of both the Champions League and the Europa League. In particular the authors focussed on the penultimate match of the group to determine the team's level of performance in the final match day when they were guaranteed first or last place. Page and Page found that teams who were guaranteed first place in the group prior to their final match would perform poorly compared to their other matches. This may be due to a number of reasons such as managerial changes to the starting team in order to rest players or lack of motivation by the players. The authors also noted that teams who were scheduled to play their last group match against the top team of the group were at an unfair advantage as the lower placed team had a higher chance to win the match. This finding strengthens Schembri et al.'s (2010) conclusion that there was a greater potential for match fixing if the last group match was between the first ranked team and the second ranked team. On the other hand Page and Page found that teams who were assured of finishing last would perform much better than their previous group matches. As Page and Page explain, the last placed team may be motivated to finish the group stage with a win as there is less or no pressure on the team to obtain a result. Teams may also choose to play for pride.

In the 2012 Summer Olympics held in London, there were a few speculations of match fixing. The aim of this paper was to examine any potential match fixing incidents at the London Olympics and assess whether match scheduling could minimise match fixing in the future.

2. METHODS

The results and schedules for the following events: badminton, football, and basketball were obtained from the official London 2012 summer Olympics website (London 2012 Olympics, 2012).

The rankings of each team prior to the start of the Olympic Games were obtained from the websites of the respective sport's governing bodies. Teams within each group were given a group rank based on their overall rank.

To examine the match scheduling within each sporting event, the six different schedules established from Schembri et al.'s (2010) study were used. In their study Schembri et al. investigated potential match fixing at the FIFA world cup. As each group comprised of four teams, they found that there was a total of six different match schedules that could occur within the group stages. Table 1 (Schembri et al., 2010) was used to demonstrate the six match schedules.

Table 1. The six possible schedules during the group phase of the FIFA World Cup.

	Group Ranking			
Schedule	1[a]	2	3	4
1	2[b], 3, 4	1, 4, 3	4, 1, 2	3, 1, 2
2	3, 4, 2	4, 3, 1	1, 2, 4	2, 1, 3
3	2, 4, 3	1, 3, 4	4, 1, 2	3, 1, 2
4	3, 2, 4	4, 1, 3	1, 4, 2	2, 3, 1
5	4, 3, 2	3, 4, 1	2, 1, 4	1, 2, 3
6	4, 2, 3	3, 1, 4	2, 4, 1	1, 3, 2

[a] Indicates the ranking of the team within the group
[b] Indicates that the highest ranked team played the third ranked team in their first match

As seen in Table 1, the 1st match in schedule 1 and 3 saw the 1st ranked team of the group play against the 2nd ranked team. Meanwhile the last match in both schedule 2 and 5 was between the 1st and 2nd ranked team.

3. RESULTS

In the 2012 London Olympics, there were a few well publicised controversial instances of match fixing which included, the women's badminton, women's football and the men's basketball. As these sports had a large media coverage due to their controversial matches, the match scheduling used for the badminton, football and basketball were analysed.

Badminton

Table 2. Group A of the women's double badminton prior to the final group match

			Games			Points		
Team	**W**	**L**	**W**	**L**	**Dif**	**W**	**L**	**Dif**
China[a]	2	0	4	0	4	84	33	51
Korea[a]	2	0	4	0	4	86	55	31
Russia	0	2	0	4	-4	54	86	-32
Canada	0	2	0	4	-4	34	84	-50

[a]Disqualified

Based on Schembri et al.'s (2010) possible match schedules, Group A in the women's doubles badminton tournament used schedule 5. This schedule involved the top two ranked teams playing against each other in the last group match. Therefore the last match in this group would be between the 1st ranked Chinese team of Yu Yang and Wang Xiaoli against the 2nd ranked Korean pairs of Jung Kyung Eun and Kim Ha Na. As seen in Table 2, prior to the final match between Korea and China both teams were guaranteed to qualify to the quarterfinals.

Table 3. Group A of the women's double badminton prior to the final group match if Schedule 1 was utilised

			Games			Points		
Team	**W**	**L**	**W**	**L**	**Dif**	**W**	**L**	**Dif**
Korea[a]	2	0	4	0	4	84	41	43
China[a]	1	1	2	2	0	67	57	10
Russia	1	1	2	2	0	57	60	-3
Canada	0	2	0	0	-4	34	84	-50

Table 3 demonstrates the group standings before the final group matches using schedule 1. Schedule 1 involves the top two ranked teams playing each other in the 1st round (China vs Korea). The 2nd round match would be between China and the 3rd ranked team (Russia). The last match for the 1st ranked team would be against the 4th ranked team (China vs Canada). The other final match of the group would between Korea vs Russia. The use of this schedule seems to create less chances of match fixing as Korea is not guaranteed to advance to the quarter finals as a win to China and Russia could see them miss out.

Table 4. Group C of the women's double badminton prior to the final group match

			Games			Points		
Team	**W**	**L**	**W**	**L**	**Dif**	**W**	**L**	**Dif**
Korea[a]	2	0	4	0	4	84	41	43
Indonesia[a]	2	0	4	1	3	104	74	30
Australia[b]	1	2	3	4	-1	114	120	-6
South Africa[b]	0	3	0	6	-6	59	126	-67

[a]Disqualified
[b]Played 3 matches instead of 2 due to fixture used at the Olympics

Group C of the women's double badminton used the 5th schedule, however as seen in Table 4 both Australia and South Africa had played all their matches prior to the final round due to the fixtures used. With two teams completing their matches the day before the final match between Korea and Indonesia, both Korea and Indonesia knew that they had qualified for the quarterfinals.
The last match of this group was between the 1st ranked Koreans (Ha Jung Eun and Kim Min Jung) and 2nd ranked Indonesians (Greysia Polii and Meiliana Jauhari). The winner would finish in 1st place.

Table 5. Group C of the women's double badminton prior to the final group match if schedule 1 was utilised

			Games			Points		
Team	**W**	**L**	**W**	**L**	**Dif**	**W**	**L**	**Dif**
Korea	2	0	4	1	3	102	73	29
Australia	1	1	2	2	0	68	58	10
Indonesia	1	1	2	2	0	89	88	1
South Africa	0	0	0	2	-2	44	84	-40

If, however, schedule 1 was used rather that schedule 5 the group would be very even with 3 teams still able to qualify for the quarterfinals. In Table 5 it can be seen that Korea, Australia and Indonesia all had a chance to advance. The final matches in this schedule would be Korea vs South Africa and Indonesia vs Australia. Therefore the match between the 2nd and 3rd ranked teams in this group (Indonesia vs Australia) would determine which team would proceed to the knock out stages of the tournament.

Football

Table 6. Group F of the women's football prior to the final group match

Team	W	D	L	GF	GA	GD	P
Sweden	1	1	0	4	1	3	4
Japan	1	1	0	2	1	1	4
Canada	1	0	1	4	2	2	3
South Africa	0	0	2	1	7	-6	0

Schedule 4 was used in Group F of the women's football tournament. This schedule involved the 1st ranked team playing the 3rd ranked team in the 1st round of matches. In this case, Japan against Canada. The 2nd match for Japan was against the 2nd ranked team (Sweden). This left the last match of the group between Japan and South Africa (4th ranked).

The women's football tournament comprised of three groups with the top two teams and the two best 3rd place teams advancing to the quarterfinals. Teams were awarded three points for a win, one point for a draw and no points for a loss. Therefore, prior to the final group matches, Sweden and Japan were both guaranteed a quarterfinals' place. However all teams in the group had a potential to qualify for the quarterfinal stage. A victory for South Africa against Japan could potentially elevate them to third place if results went their way (for example, Canada losing to Sweden by a large margin). Meanwhile a victory for Canada against Sweden could result in them finishing in 1st position, depending on Japan's result against South Africa (a loss or draw to Japan would benefit Canada).

Table 7. Group F of the women's football prior to the final group match if Schedule 5 was utilised

Team	W	D	L	GF	GA	GD	P
Sweden	1	1	0	6	3	3	4
Japan	1	1	0	2	1	1	4
Canada	0	1	1	3	4	-1	1
South Africa	0	1	1	1	4	-3	1

Table 7 displays the group standings before the final match day. Unlike schedule 4, this schedule results in a group were any team is still able to qualify for the quarterfinal stages of the tournament. The final matches are between the top two ranked teams (Japan vs Sweden) and the bottom two ranked teams (Canada vs South Afica). The winner of the Japan and Sweden game would finish the group stage on top. Meanwhile a win for Canada or South Africa could see either side qualify as second or best 3rd team (depending on the goal difference).

Basketball

Table 8. Pool B of the men's basketball prior to the final group match

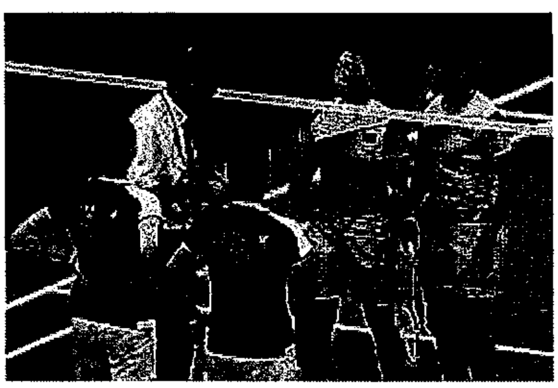
Team	W	L	GF	GA	GD	P
Brazil	3	1	314	267	47	7
Russia	3	1	320	277	43	7
Spain	3	1	332	306	26	7
Australia	2	2	328	293	35	6
Great Britain	0	4	325	325	-35	4
China	0	4	349	349	-94	4

Table 8 reveals that Brazil, Russia and Spain were all guaranteed a quarterfinal place. However the winner of Group B would be decided in the last group matches. The final set of matches involved Australia vs Russia, Great Britain vs China and Spain vs Brazil. In the 1st set of matches Russia had a chance to extend their lead at the top of the table with a win, meanwhile Australia could potentially finish 3rd with a win (this would be decided by a head to head record between the teams with equal points). At the Olympics, teams were awarded two points for a win and one point for a loss. Therefore the second match between Great Britain and China was a battle for 5th place in the group. The last match of the group between Brazil and Spain was the top of the table clash, with the winner potentially finishing top of the table if the results went their way (Australia beating Russia).

4. DISCUSSION

Badminton

In the 2012 Summer Olympics held in London, there were a few controversial instances of match fixing. In particular the Women's badminton doubles, where four teams were disqualified by the Badminton World Federation (BWF). These players breached Sections 4.5 and 4.16 of the Players' Code of Conduct by "not using one's best efforts to win a match" and "conducting oneself in a manner that is clearly abusive or detrimental to the sport" (Badminton World Federation, 2012a).



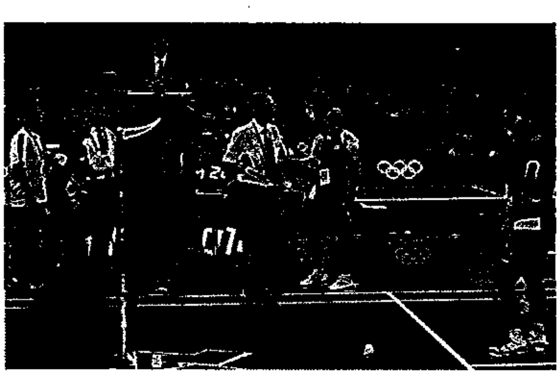


Figure 1: The players involved were China's world champions Wang Xiaoli and Yu Yang, Indonesia's Greysia Polii and Meiliana Jauhari and two South Korean pairs – Jung Kyung Eun and Kim Ha Na, and Ha Jung Eun and Kim Min Jung. All are receiving warnings of a black card.

In Group A the top two ranked teams, China and Korea had already both qualified for the quarter finals with two wins. Therefore the last match between the two nations would determine who would top the standings.
Prior to the final group match in group A between China and Korea, the match between China and Denmark took place where the Danish pair recorded an upset victory to beat the 2nd seeds (Badminton World Federation, 2012b). This result meant that the Danish, Chinese and Japanese all ended on two wins, however Denmark and China both qualified for the knock-out round due to a superior game difference. This result meant that a win for the number one ranked Chinese pair against Korea in group A could potentially see anall Chinese semi-final. However in the controversial match neither the Chinese or Korean players wanted to win the match. Both teams played in an uncharacteristic manner with bad serving and returning. The match lasted only 23 minutes with Korea's Jung Kyung Eun and Kim Ha Na eventually winning 21-14, 21-11. By losing their last group match the Chinese pair of Yu Yang and Wang Xiaoli avoided a semi-final match against their compatriots Tian Qing and Zhao Yunlei. This meant that both Chinese pairs would meet in the final, if they had won their respective quarter-final and semi-final matches.

The other two teams disqualified for breaching the player's code of conduct were the Korean and Indonesian pairs, who also played against each other in the last group match. As seen in the other controversial match, both teams in group C had already qualified for the quarterfinals with two wins each. Prior to the match taking place both teams knew that a win would lead to a quarterfinal match against the number one seeded Chinese pair and a loss would mean a match against the 8th seeded Korean team. As seen in the earlier group A match, neither the Korean and Indonesian teams wanted to win. It seemed that both teams wanted to avoid playing the Chinese in the quarterfinals. In the end Korea won the match against Indonesia 18-21 21-14 21-12, and therefore would play against China in the quarterfinal.

The match scheduling for the Women's and Men's badminton was pre-determined with the highest ranked pair playing against the 4th ranked pair the first round of matches. The 2nd round of matches involved the 1st ranked team playing the 2nd ranked team.

A change in match schedule can be used in order to minimise the chances of match fixing in the future. When the 1st ranked and 2nd ranked pair play in the 1st round of matches, there are less opportunities for match fixing as both teams would prefer to start the tournament off with a win. If this match schedule was used in the two controversial groups then 3 teams had the potential to qualify for the knockout stages compared to the original match schedule where the two quarterfinalists were known prior to the last group match.

As seen in the results, the group standings prior to the final match was vividly different in both groups when schedule 1 was utilised rather than schedule 5. In group A all the teams could potentially qualify for the knockout rounds apart from Canada. Therefore this schedule would help minimise match fixing as the top ranked team, China, is required to win their final match against the 2nd ranked Koreans as a loss could see them drop to 3rd place (if Russia beat Canada in the other group match).

A change in schedule could also benefit group C to minimise match fixing. Through the use of schedule 5 the group standings also change and three teams could advance to the quarterfinals. The final group matches in schedule 5 are between Korea vs Indonesia and Australia vs South Africa. A win by Indonesia and Australia could see Korea missing out of the quarterfinals (in this case the top two teams based on game or point difference).

Due to the disqualification of the four teams; Russia, Canada, Australia and South Africa all qualified for the quarterfinals. The Chinese pair of Tian Qing and Zhao Yunlei won the gold medal match against the Japanese pair of Mizuki Fujii and Reika Kakiiwa (21 - 10 25 – 23). Third place went to Valeria Sorokina and Nina Vislova of Russia.

Football

Speculation of match fixing also occurred in the women's football tournament at the 2012 Olympics. In the last group F match Japan, the reigning world cup champions, played out an unconvincing scoreless draw with South Africa who was the lowest ranked team in the whole tournament. Prior to kick-off at Millennium Stadium in Cardiff, the Japanese knew that a win against South Africa would see them top the group and be required to travel Glasgow, Scotland for their quarterfinal match. However if the Japanese drew their match, they would stay in Cardiff, Wales and avoid the long trip to Glasgow. With Japan already qualified for the quarter-finals, Japanese coach Norio Sasaki started the game against South Africa with only four regular starters. Despite dominating possession (65% vs 35%) the Japanese failed to score. After the match the Japanese coach had this to say about his team's performance: "I feel sorry we couldn't show a respectable game, but it's my responsibility, not the players', why the game was like that. It was important for us not to move to Glasgow" (National Broadcasting Company, 2012a).

After their draw, Japan had to wait a couple hours to discover the team they would be facing in the quarterfinals (the loser between Great Britain and Brazil). In an unexpected result, Great Britain beat Brazil 1-0 and thus ensured top spot in group E. Therefore the Japanese team would have to play Brazil who was the 3rd ranked team in the world in the quarterfinals. After the group matches were finalised, Japan was moved to the opposite side of the draw to the USA meaning that the two 2011 world cup finalists could only meet in the final. The Japanese ended up beating Brazil 2-0 and also beat France in the semi-final 2-1, resulting in a final against the number one ranked Americans. The Americans went on to win the tournament 2-1.

In order to minimise match fixing in the FIFA world cup, Schembri et al. suggested that the top two ranked teams in each group should play against each other in the final group match (schedule 2 and 5). As seen in the results, the use of schedule 5 changed the dynamics of the group. Unlike schedule 4, the schedule used at the Olympics, the group was still undecided with either Canada or South Africa able to qualify for the quarter finals.

Unlike the four badminton pairs who were disqualified for trying not to win a game in order to play an easier opponent in the knockout rounds, the Japanese football team intentionally drew their game against South Africa to avoid travelling to another country for their quarterfinals.

Basketball

There was yet another controversial incident in the men's basketball match between Spain and Brazil. With both Spain and Brazil guaranteed a quarterfinal berth, the last pool B match would determine which team would finish in 2nd place. A win to either team would result in a semi-final against the number one ranked Americans, while the loser would only meet the Americans in the final. Brazil won the match 88-82 and secured second place in pool B behind Russia, meaning a potential semi-final match against the USA. Despite leading 66-57 at the end of the 3rd quarter, the Spanish failed to keep their lead. With this loss Spain guaranteed that they would not play the USA until the final, assuming both Spain and USA would win their quarterfinal and semi-finals matches.

After the match Brazilian basketball player Guilherme Giovannoni said that the Brazilian "team played very, very hard" and "we absolutely played to win." When asked if Spain played to win, Guilherme replied "You will have to ask them" (August 6, 2012b). Meanwhile the Spanish players denied any match fixing: "We always play to win," said Jose Caldera. "We're in the Olympics and you have to try to win every game." National Broadcasting Company (August 6, 2012b).

As the runner up in pool B, Brazil played against Argentina where they lost 77-82. Meanwhile, Spain won their quarterfinal match against France (66-59). Following this victory was a semifinal win against Russia (67-59). The gold medal match was therefore to be decided between the USA and Spain, the number one and number two ranked teams respectively. The American's won the match 107–100 and thus defended their Olympic gold medal from Beijing 2008.

5. CONCLUSION

Notably, match fixing at the Olympics was evident in a variety of sports, with the expulsion of four badminton teams being the most significant event. A series of applications of existing schedules would have provided these sports less of an opportunity for match fixing events. The use of better scheduling programs presents another option for tournament organisers to reduce the risk of match fixing.

6. REFERENCES

Australian Government (June 10, 2011) National policy on match-fixing in sport. Sourced from http://www.regional.gov.au/sport/programs/match_fixing.aspx

Badminton World Federation (2012a). Sourced from http://www.bwfbadminton.org/news_item.aspx?id=65297

Badminton World Federation (2012b). Sourced from http://www.bwfbadminton.org/news_item.aspx?id=65366

London 2012 Olympics (2012) Sourced from http://www.london2012.com/

McLaren, R. H. (2008). Corruption: Its impact on fair play, *Marquette Sports Law Review, 19*(1), 15-38.

National Broadcasting Company (August 1, 2012a). Japan draws South Africa, but that was the plan. Sourced from http://www.nbcolympics.com/news-blogs/soccer/south-africa-holds-japan-to-draw.html

National Broadcasting Company (August 6, 2012b). Harder to spot tankers in hoops. Sourced from http://www.nbcolympics.com/news-blogs/basketball/harder-to-spot-tankers-in-hoops.html

Page, L., & Page, K. (2009). Stakes and motivation in tournaments: Playing when there is nothing to play for but pride. *Economic Analysis & Policy, 39* (3), 455-464.

Preston, I., & Szymanski S. (2003). Cheating in contests, *Oxford Review of Economic Policy, 19*(4), 612-624.

Schembri, A. J., Golkhandan, S. R., & Bedford, A. (2010). 'Re-scheduling to minimize the potential for match fixing at the FIFA World Cup', *Proceedings of the 10th Australasian Conference on Mathematics & Computers in Sport*, 165-172.

HEPTATHLON: WHICH EVENT IS THE BEST PREDICTOR OF TOTAL POINTS?

Ian Heazlewood[a, b]

[a] *Charles Darwin University*
[b] *Corresponding author: ian.heazlewood@cdu.edu.au*

Abstract

The heptathlon event is conducted over two consecutive days and conducted in the flowing order. Day 1, the 100m hurdles, high jump, shot put and 200m. Day 2, the long jump, javelin throw and 800m. The individual event performances are converted to points using IAAF heptathlon scoring tables and then these points are summed to assess rank performance and award athlete overall event place. The multivariate statistical methods of structural equation modelling (Heazlewood, 2011) and factor analysis (Heazlewood, 2008) have been applied to understand the statistical interrelationships between the seven events, however the importance of each event and sets of events that contribute to and predict the total points require further investigation as research thus far has not addressed this research question. The data base to derive the regression solutions were based on the International Association of Athletic Federations (abbreviated as IAAF) top ranked 173 women heptathletes for the 2010 competition year (IAAF, 2010). The results indicated the long jump and shot put predict a significant 59% of the explained variance. A three event model with the inclusion of the 100m hurdles adds an additional 11.7% making 70.7 % explained variance. The additional events of high jump, javelin, 800m and 200m contribute 27% to 97.7%. Al events contribute significantly as the final points are a linear additive score. The significant events predicting total points were long jump, shot put and 100m hurdles suggesting additional training focussed on these events will contribute to total points achieved in the women's heptathlon.

Keywords: Coaching, heptathlon, multiple regression, prediction

1. INTRODUCTION

The heptathlon event is conducted over two consecutive days and conducted in the flowing order. Day 1, the 100m hurdles, high jump, shot put and 200m. Day 2, the long jump, javelin throw and 800m. The individual event performances are converted to points using IAAF heptathlon scoring tables and then these points are summed to assess rank performance and award athlete overall event place. Identified published research indicated that most heptathlon articles refer to performance characteristics of high performance heptathletes, as well as, some conceptualisations concerning the underpinning motor fitness factors that contribute to each event. These conceptual models have been proposed based on exercise physiological and underpinning motor fitness constructs for each event, such as by Hancock (1987), Mackenzie (2007), Marra (1985), Sarponov (1982), and Telfer (1988). One these more detailed conceptual models postulated by Mackenzie (2007) is presented in table 1. In this table Mackenzie (2007) has attempted to assign relative conceptual weights for each event with constructs of aerobic endurance, gross strength skill, relative strength, running speed, mobility, explosive strength-power, speed endurance and strength endurance that are believed to underpin each event. Research by Heazlewood (2010) using decathlon World ranked athletes indicated that the shot put one of the weakest events as a percentage of World record was the most predictive event for total decathlon points.

Event	Aerobic Endurance	Gross strength	Skill	Relative Strength	Running Speed	Mobility	Explosive Strength	Speed Endurance	Strength Endurance
100m Hurdles	-	Med	High	High	High	High	High	Med	-
High jump	-	Low	High	High	High	High	High	-	-
Shot put	-	High	High	Med	Low	Med	High	-	-
200m	Low	Med	Med	High	High	High	High	High	High
Long Jump	-	Low	High	High	High	High	High	-	-
Javelin	-	Med	High	High	Low	High	High	-	-
800m	High	-	Low	Low	Med	Low	-	-	High

Table. 1. Hypothesized Motor Fitness Constructs of Mackenzie (2007) that Underpin the Seven Events in the Women's Heptathlon.

In the sport of heptathlon the multivariate statistical methods of structural equation modelling (Heazlewood, 2011) and factor analysis (Heazlewood, 2008) have been applied to understand the statistical interrelationships between the seven events, however the importance of each event and sets of events that contribute to and predict the total points were not addressed. The aim of the research was directed towards evaluating which event or subset of events is/are the best predictor of total points and may indicate which event or events may require additional training focus based on the predictive importance for overall heptathlon performance.

2. METHODS

The data base to derive the regression solutions were based on the International Association of Athletic Federations (abbreviated as IAAF) top ranked 173 women heptathletes for the 2010 competition year (IAAF, 2010). In the data set used in this research the track events were measured to the nearest 0.01 second and field events to the nearest 0.01 metre and according to IAAF rules. The initial relationship of the individual events with total heptathlon points were evaluated by Pearson product moment bivariate correlations and to assess if any significant exist. Based on significant bivariate relationships multiple linear regression was the applied to develop a multivariate linear additive model of the seven events as the independent/predictor variables and the total points as the dependent variables. The stepwise method of multiple regression was applied to evaluate a statistical solution to deriving a model.
In more detail regression analysis represents a linear additive statistical modelling approach where:

- Multiple regression (Hair et al., 2010; Garson, 2011) can establish if a set of independent variables (seven heptathlon events) explain a significant proportion of the variance in a dependent variable at a significant level through a significance test of R^2 and change in R^2. In this context total heptathlon points based on the seven events.
- Establish the relative predictive importance or influence of the independent variables by comparing standardized beta weights and collectively by the variance explained in the model (R^2).
- Power terms can be added as independent variables to explore curvilinear effects, if they exist. Cross-product or interaction terms can be added as independent variables to explore interaction effects.
- Test the significance of difference of two R squares (R^2) to determine if adding an independent variable such as one of seven events to the model, helps significantly.
  - A subset of significant variables can be identified that explain the majority of the explained variance.
  - Stepwise method in regression derives an order of statistical importance based on inclusion and exclusion criteria, which

iterates until no omitted variable are included on statistical evidence.

3. RESULTS

Table 2 indicates the means, standard deviations, current world records and the heptathlon mean for each seven event as a percentage of the current world record.

Event	Mean	Std. Deviation	World Record	% Record
Hurdles (s)	14.00	.37	12.21	87.21
High jump (m)	1.74	.07	2.09	83.44
Shot (m)	12.68	1.15	22.63	56.07
200m (s)	24.98	1.05	21.34	85.43
Long jump (m)	6.03	.25	7.52	80.19
Javelin (m)	41.61	5.37	72.28	57.57
800m (s)	137.73	5.14	113.28	82.23

Table 2. Descriptive Statistics for each Heptathlon Event.

It is interesting to observe the throwing events, such as the shot put (56.07%) and javelin (57.57%) have the lowest percentage of current world record, whereas the 100m hurdles (87.21%) and the 200m (85.43%) have the highest percentage of current world record and the remaining events of high jump, long jump and 800m are all above 80% of world record standard.

Table 3 indicates the Pearson product moment bivariate correlations between the seven events and total points. It can be noted the order of significant correlation of each event with total points is long jump (.590), shot put (.549), hurdles (-.515), high jump (.510), 80m (-.395), 200m (-.350) and javelin (.154). All events except the javelin ($p<.05$) are statistically significant at $p<.01$at the bivariate level. Javelin correlation is small in magnitude and the coefficient of determination is very small ($r^2 = .024$ or 2.4%), however significant due to the large sample size. This indicates in the bivariate context that this event appears to be the least influential.

Event		points	hurdles	high jump	shot	200m	long jump	javelin	800m
points	Pearson Correlation	1	-.515**	.510**	.549**	-.350**	.590**	.154*	-.395**
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.043	.000
	N	173	173	173	173	173	173	173	173
hurdles	Pearson Correlation	-.515**	1	-.080	-.127	.386**	-.223**	-.034	.159*
	Sig. (2-tailed)	.000		.298	.095	.000	.003	.656	.037
	N	173	173	173	173	173	173	173	173
high jump	Pearson Correlation	.510**	-.080	1	.109	-.097	.332**	.032	-.046
	Sig. (2-tailed)	.000	.298		.155	.205	.000	.672	.544
	N	173	173	173	173	173	173	173	173
shot	Pearson Correlation	.549**	-.127	.109	1	.005	.101	.093	-.102
	Sig. (2-tailed)	.000	.095	.155		.947	.186	.224	.183
	N	173	173	173	173	173	173	173	173
200m	Pearson Correlation	-.350**	.386**	-.097	.005	1	-.255**	.048	.194*
	Sig. (2-tailed)	.000	.000	.205	.947		.001	.530	.011
	N	173	173	173	173	173	173	173	173
long jump	Pearson Correlation	.590**	-.223**	.332**	.101	-.255**	1	.030	-.061
	Sig. (2-tailed)	.000	.003	.000	.186	.001		.693	.429
	N	173	173	173	173	173	173	173	173
javelin	Pearson Correlation	.154*	-.034	.032	.093	.048	.030	1	-.088
	Sig. (2-tailed)	.043	.656	.672	.224	.530	.693		.248
	N	173	173	173	173	173	173	173	173
800m	Pearson Correlation	-.395**	.159*	-.046	-.102	.194*	-.061	-.088	1
	Sig. (2-tailed)	.000	.037	.544	.183	.011	.429	.248	
	N	173	173	173	173	173	173	173	173

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

Table 3. Correlations for Seven Events and Total Points.

Model Steps	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.590[a]	.348	.344	212.68992	.348	91.322	1	171	.000
2	.768[b]	.590	.585	169.33782	.242	100.270	1	170	.000
3	.841[c]	.707	.702	143.62563	.117	67.316	1	169	.000
4	.891[d]	.794	.789	120.75407	.087	71.082	1	168	.000
5	.942[e]	.887	.883	89.82838	.093	136.589	1	167	.000
6	.983[f]	.966	.965	49.01527	.080	394.871	1	166	.000
7	.988[g]	.977	.976	40.60619	.011	76.882	1	165	.000

Table 4. Stepwise Multiple Regression Model Summary.

a. Predictors: (Constant), long jump.
b. Predictors: (Constant), long jump, shot.
c. Predictors: (Constant), long jump, shot, hurdles.
d. Predictors: (Constant), long jump, shot, hurdles, high jump.
e. Predictors: (Constant), long jump, shot, hurdles, high jump, javelin.
f. Predictors: (Constant), long jump, shot, hurdles, high jump, javelin, 800m.
g. Predictors: (Constant), long jump, shot, hurdles, high jump, javelin, 800m, 200m.

Dependent Variable: Total Heptathlon points.

From the preceding results it can be observed that the long jump and shot put predict a significant 59% of the explained variance. A three event model with the inclusion of the 100m hurdles adds an additional 11.7% making 70.7 % explained variance. The additional events of high jump, javelin, 800m and 200m contribute 27% to 97.7%. Al events contribute significantly as the final points are a linear additive score. Negative correlations occur as lower times equates with higher performance.

Model	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
(Constant)	5354.010	186.120		28.766	.000
Long jump	339.583	13.385	.330	25.371	.000
shot	74.072	2.827	.326	26.198	.000
hurdles	-181.547	9.093	-.261	-19.965	.000
High jump	1173.955	44.907	.328	26.142	.000
javelin	17.042	.605	.349	28.190	.000
800m	-13.955	.621	-.273	-22.481	.000
200m	-29.209	3.331	-.116	-8.768	.000

a. Dependent Variable: Total Heptathlon points.

Table 5. Unstandardised, Standardised Beta Coefficients and Equation Constant for the Multiple Regression Equation. The order from long jump to 200m represents the order of inclusion in the stepwise regression solution.

The predictive equation built from the coefficients displayed in table 5 is:

Total points = (339)long jump + 74(shot put) – (181)100m hurdles + (1173)high jump + (17)javelin – 14(800m) – (29)200m + 5354. **(1)**

It is interesting to note the shot put is the second most significant event in predicting total points.

4. DISCUSSION

The regression equation predicts essentially all of the explained variance in the model 97.7% and all the events of course are predictive and statistically significant as they contribute to the final total points. The shot put is the second most significant event in predicting total points and similar to the men's decathlon where the shot put was the most significant predictor of the total decathlon points (Heazlewood, 2010) even though the event was the poorest in terms of performance compared to world record standard.

In term of training high performance heptathletes might have more to gain re total points by improving this event. Although the javelin has a similarly low percentage of World record standard it was not as predictive as the shot put and at the bivariate level the overlap with total points was very small. All running (800m) sprinting (200m and 100m hurdles) and jumping events (high jump and long jump) were in excess of 80% of world records, however room to move in terms of points achieved per these events might be limited, as greater potential improvement is within the throwing events.

The previous research of factor structure of the heptathlon (Heazlewood, 2008, 2011) indicates the jumps, throws and sprints load on orthogonal factors, as well as displaying relatively low interrelationship at a bivariate correlation level emphasising the individual contribution of each event to total points.

In terms of training the three important events, which explain the majority of the variance are long jump, shot put and 100m hurdles and in this order. An emphasis of training and improving performance in these three events should translate to increased total points achieved in competition based on the findings of stepwise multiple regression.

5. CONCLUSIONS

The predictor of the total points in order of importance based on the multiple regression stepwise method of Predictors: (Constant), long jump, shot, hurdles, high jump, javelin, 800m and 200m. The three events which explain and predict the majority of explained variance were the long jump, shot put and 100m hurdles, which suggest an additional training emphasis on these events to enhance performance and total points achieved in the heptathlon. The next step in the research process is to take other IAAF years in terms of heptathlete rankings to assess if the predictive model is replicated across different data sets and test the robustness of the model.

References

Hair, J., Black, W., Babin, B. & Anderson, R. (2010). Multivariate Data Analysis. Sydney: Pearson.

Hancock, T. (1987). Problem confronting coaches in multiple events. Modern Athlete and Coach. 25 (3), pp.36- 38.

Heazlewood, T. (2008). Factor Structure of the Women's Heptathlon: Implications for Training. Proceedings of First Joint International Pre-Olympic Conference on Sports Sciences and Sports Engineering. Volume I: Computer Science in Sport. Ed. Y. Jiang, A. Baca & H Zhang. Nanjing, China, August 4-7, 2008. pp. 283-288.

Heazlewood, I. (2010). Predicting Performance in High Performance Decathletes Based on Multiple Linear Regression and Discriminant Analysis. In Korusuz, F., Ertan, F. & Tsolakidis, E. (Eds). Refereed Book of Abstracts. 15th Annual Congress of the European College of Sport Science. Antalya/Turkey, 23-26 June, 2010. P. 51-52.

Heazlewood, I.T. (2011). Factor structure of the women's heptathlon: applications of traditional factor analysis and structural equation modelling. Theories and Applications: The International Edition. March 2011, Volume 1, No. 1, 114 – 125.

Garson, D. (2011). Multiple Regression. Statnotes: Topics in Multivariate Analysis. Retrieved 6/1/2011 from http://faculty.chass.ncsu.edu/garson/pa765/statnote.htm.

Mackenzie, B. (2007). Heptathlon. Sports Coach. Retrieved 5/1/2007 from http://www.brianmac.co.uk/hepth/,2007.

Marra, M. (1985). Decathlon hints. Modern Athlete and Coach, 2 (4), pp.15-18.

Sarponov, E. (1982). Looking at the future of the heptathlon, Modern Athlete and Coach, 20 (1), pp.39-40.

Telfer, P. (1988). The javelin in the heptathlon, Modern Athlete and Coach, 26 (3), pp.32-34.

Methods of Prediction in Track Athletics

Yu Cen Kang [1,2], Anthony Bedford [1] and Adrian J Schembri [1]

[1]*School of Mathematical & Geospatial Sciences, RMIT University, Melbourne, Australia*
[2] *Corresponding author: yucen.kang@rmit.edu.au*

Abstract

In this research we look at regression models for track events and compare this to simulation fits. We find great variability in the differing approaches, and note the variability in results with certain events proving far more accurate than others to predict. We compare top 5, top 10 and winner times by both athlete and events. We note the group based results yielding tighter prediction intervals, and that regression is a dangerous method when considering out-of-sample prediction.

Keywords: Simulation, regression, athletics, track.

1. INTRODUCTION

Mathematicians have attempted to predict the winning running time of each tournament in specific year, including the world records in track (or/and field) events since 1900s. In 1906, Kennelly examined the relationships between velocity and distance for various track events on a log-log scale and showed that a linear relationship between them was stable over all events. Afterwards, the evolution and prediction of winning running time of different tournaments has been extensively investigated by mathematicians (and later statisticians), psychometricians, physiologists and biomechanists (Chatterjee & Chatterjee 1982; Deakin, 1967). During the last few decades, more efficient prediction models have been developed and made more accurate. Under the assumption that the performances in track (or/and field) events have a clear trend, linear and non-linear models have been used for evaluating and predicting the winning running time, including world records since the 1900s. Linear (Regression) Models have been used to fit two different types of data for the purposes of evaluation and prediction: 1) Running performance over multiple events in a specific tournament/year (Kennelly, 1906); and 2) Best winning running times including world records over years for a single tournament (Ryder, Carr & Herget, 1976). Non-Linear (Regression) Models have been used by many researchers to predict and evaluate winning running time in track (or/and field), including (Lucy, 1958). Since these works, the two most common models used are exponential and polynomial models. The initial non-linear model used by Lucy (1958) was similar to the exponential model:

$$T(n) = b_0 + b_1 a^n \quad (1)$$

where b_0, b_1 and a are the constants to be determined, and n is the time in years.

Literature shows that different researchers (e.g. Schutz, Carr & Halliwell, 1975; Chatterjee & Chatterjee, 1982) typically used different data in their studies – obviously due to the time of publication. One consideration we need to make is that athletes may not necessarily run out the race, or run at their optimal times each race. This does have a bearing on regression modeling, hence the desire to simulate.

In this paper, we aim to fit a variety of parametric regression models and look into the ability of these models to predict times. We then determine the differences between these times and fitted simulation data and their resultant prediction times. This shall help us with future work, looking into time split modelling of athletes by both methods.

2. METHODS

2.1 Data Collection Procedures

Eight events (100m, 200m, 400m, 800m, 1500m, 5000m, 10000m and Marathon) were analysed in this paper. These events were selected because the competitive conditions are identical for both men and women in most international tournaments. Three similar types of data were collected for the paper: the last 10 and 5 results from the top athletes in each of the specific distances who appear on the top 5 of each tournament. These data can be obtained from International Association of Athletics Federations' (IAAF) official web link: www.iaaf.org.

2.2 Statistical Models of Running Performance

There were ten different prediction models used for the track events above. In the two general descriptions below, we will present how the two simplest of the ten prediction models were used briefly.

A linear prediction model of $Time_{seconds}$ using $Tournament$ for a specific event is simply given by

$$Time_{seconds}(Tournament) = b_0 + b_1(Tournament) \quad (2)$$

where $Time_{seconds}(Tournament)$ is the running time for a specific event; $Tournament$ is the recent tournament in specific years; b_0 and b_1 are the calculated parameter estimators as seen in e.g. Ballerini & Resnick, 1985, 1987. Figure 1 displays an example of a simple linear regression for Bolt.

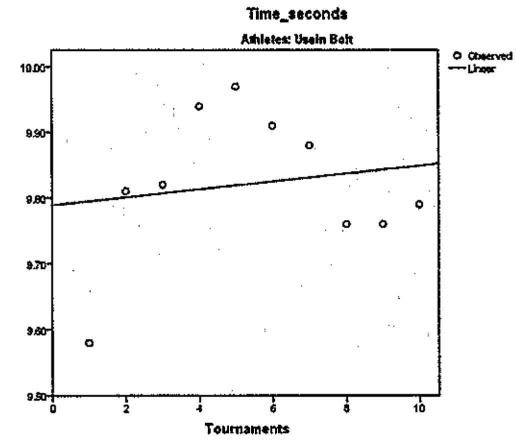


Figure 1 Tournaments and Time_seconds Linear Model for Usain Bolt

A prediction polynomial model relating $Time_{seconds}$ and $Tournament$ for a specific event is given by

$$Time_{seconds} = b_0 + b_1(Tournament) + \cdots + b_n(Tournament)^n \quad (3)$$

where $Time_{seconds}(Tournament)$ is running time for a specific event; $Tournament$ is the recent tournament in specific years; $b_0, b_1, \ldots b_n$ are the calculated parameter estimators as described in Mognoni, Lafortuna, Russo & Minetti, 1982. Figure 2 shows a polynomial fit for the same data as in Figure 1. Notably, they predict significantly different outcomes for this model.

Other prediction models of the $Time_{seconds}$ versus $Tournament$ relationship were developed separately for specific years. For developing the different tournaments' relationships, there were other specific prediction models have been used for predicting different $Time_{seconds}$'s in the paper:

$$Time_{seconds}(Tournament) = e^{b_0 + b_1 / Tournament}, \quad (4)$$

A power model :

$$Time_{seconds}(Tournament) = b_0 (Tournament)^{b_1}, \quad (5)$$

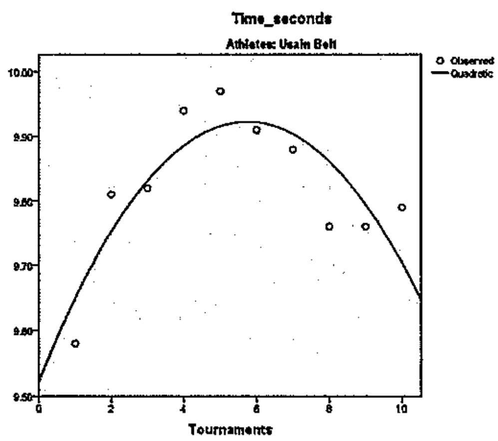


Figure 2 Time_seconds and Tournament (Polynomial Model)

The other 5 specific prediction models that have been used were Inverse, Compound, Logarithmic, Growth and Exponential; they all had similar variables with different values.

2. RESULTS

Figure 3 presents Usain Bolt's predicted winning time based upon his last 5 runs. Notably it presents a fit that increases extremely after the last run; Figure 4 presents a similar result. Figure 5 presents Shelly-Ann Fraser-Pryce's predicted winning time with her last 5 runs, it presents the prediction as a decreased time - quite extremely after the last run. Figure 6 presents Allyson Felix's predicted winning time with her last 5 runs.

Figure 7 presents the Men's 100m's predicted running time. It notably would head down after $Year = 12$; the Men's 200m's predicted running time would go higher gradually after $Year = 12$; Figure 8 presents the Men's 100m's predicted running time that would go higher slowly after $Year = 12$; and the Men's 200m's predicted running time would go extreme after $Year = 12$. Figure 8 also presents an opposing trend to Figure 7.

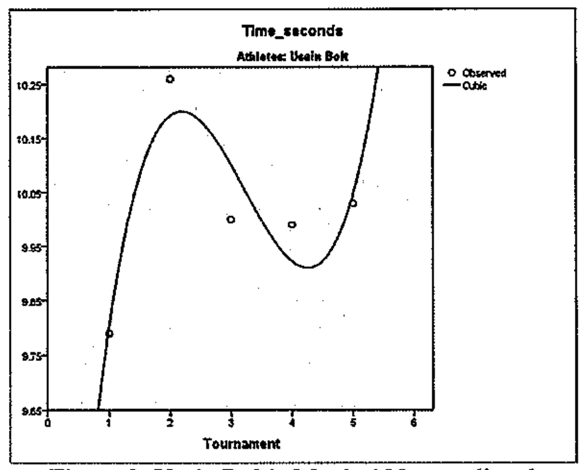


Figure 3: Usain Bolt's Men's 100m predicted winning time with last 5 runs

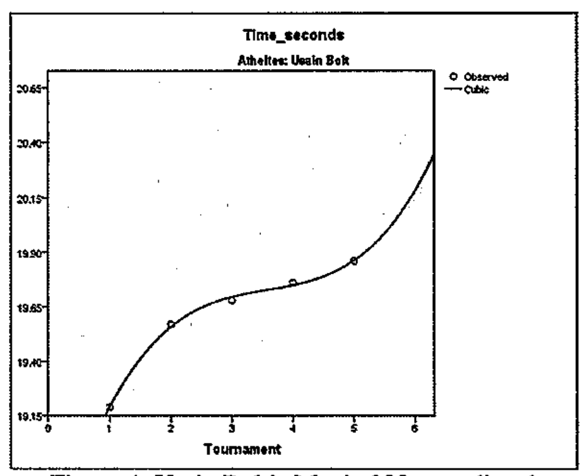


Figure 4: Usain Bolt's Men's 200m predicted winning time with last 5 runs

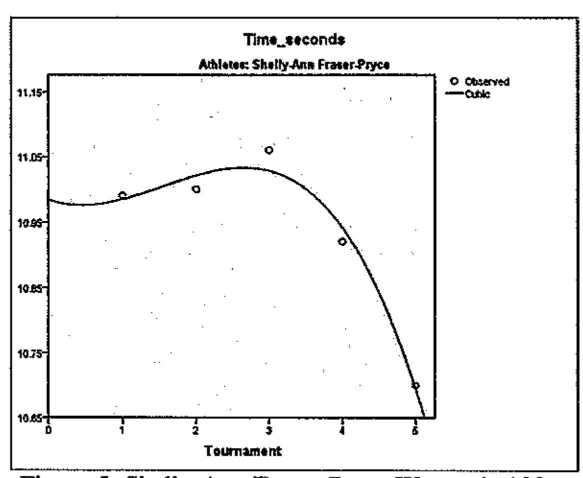


Figure 5: Shelly-Ann Fraser-Pryce Women's 100m predicted winning time with last 5 runs

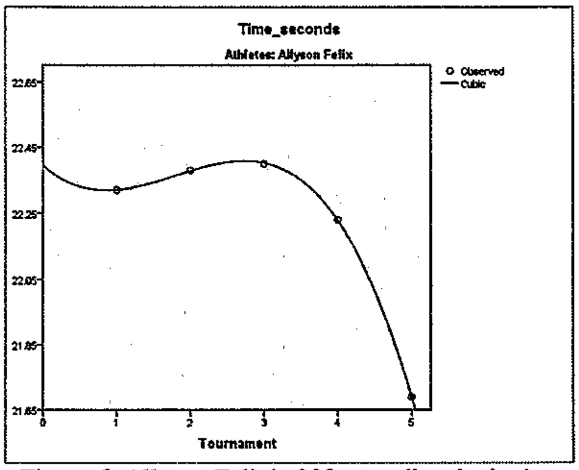


Figure 6: Allyson Felix's 200m predicted winning time with last 5 runs

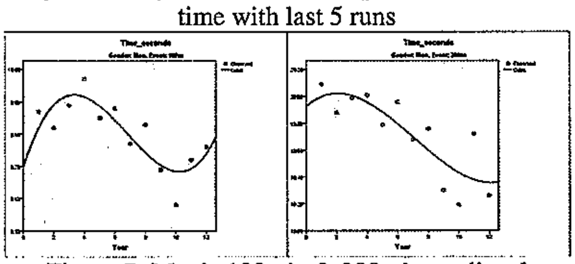


Figure 7: Men's 100m's & 200m's predicted running times with Top 5s' results per year

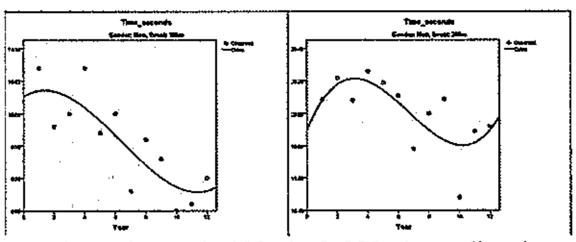


Figure 8: Men's 100m's & 200m's predicted running times with Winners' results per year

Simulation

As a comparison, we started by fitting distributions to our sets of data to determine a possible prediction set. Figure 9 presents the Logistic Distribution - optimal for predicting Usain Bolt's running time. The simulation runs up to 10,000 iterations. It yields a 95% Confidence Interval (9.670,9.959). The confidence interval tells us the lower confidence interval is 9.670s, and the upper confidence interval time is 9.959s. From the plot, we see it presents as symmetric.
Figure 10 presents a Normal Distribution fit for predicting Usain Bolt's running time - the 95% Confidence Interval, (19.138,20.383). Figure 11 presents a Logistic Distribution fit for predicting Shelly-Ann Fraser-Pryce's running time. It has a 95% Confidence Interval, (10.704,11.206). Figure 12 presents a Triangle Distribution for predicting Allyson Felix's running time. It has a 95% Confidence Interval, (21.708,22.596).

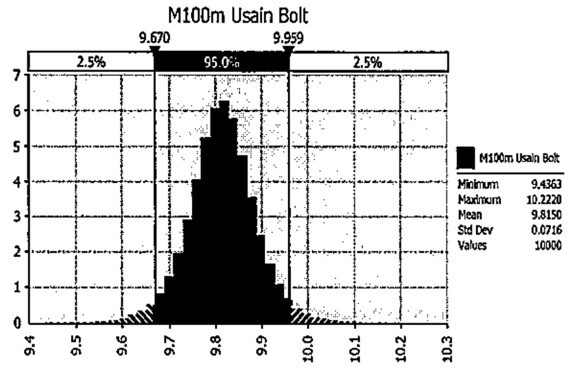


Figure 9: M100m Usain Bolt's Last 5 runs with Logistic Distribution

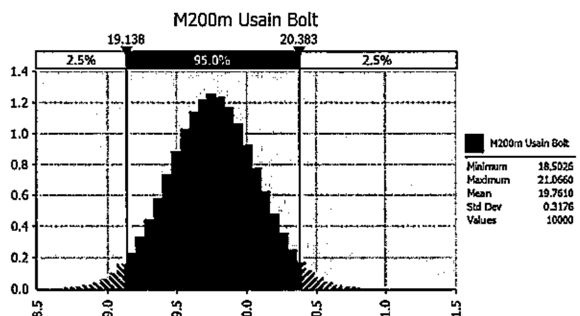


Figure 10: M200m Usain Bolt's last 10 runs with Normal Distribution

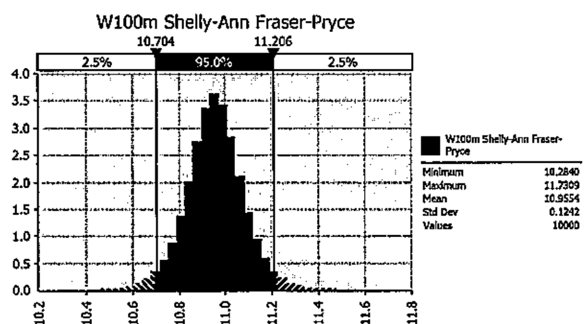


Figure 11: W100m Shelly-Ann Fraser-Pryce's Last 5 runs with Logistic Distribution

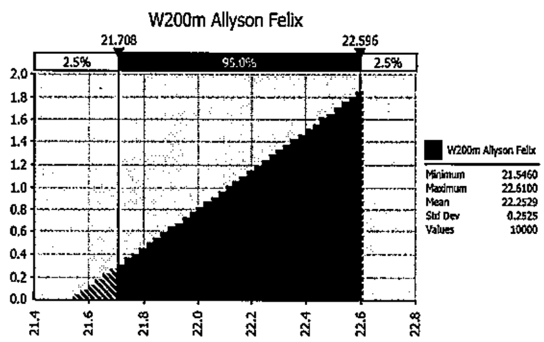


Figure 12: W200m Allyson Felix's Last 10 runs with Triangle Distribution

The prediction models in Table 1 are detailed with the best r^2 values.
We compared the predicted winning (PW) to actual 2012 Olympics Personal's (AOP) results in $Time_{seconds}$'s. This appears in the 3rd and 4th columns of Table 1; and the Predicted 95% Confidence Interval of Regression Line (95% P.I.) in the 5th column.

The only result did not present well was the M100m for Asafa Powell's AOP. This presents a prediction that seems odd.
For Veronica Campbell-Brown's W200m the predicted time yielded a 23.16s result using the following model:

$$Time_{seconds}(Tournament) = 21.714 + 0.589 \cdot (Tournament) - 0.414(Tournament)^2 + 0.009 \cdot (Tournament)^3 \quad (6)$$

We model $Tournament = 13$ for predicting the 2016 Olympics, and $Tournament = 14$ for predicting the 2020 Olympics. In the 6th and 7th columns of the table, they present the Minimum Simulation $Time_{seconds}$ and Simulation 95% Confidence Interval. Compare with the AOP and Simulation 95% C.I., we see the AOPs generally fit into the 95% C.I. For example, Usain Bolt's 100m, the AOP is 9.63s and the 95% C.I. is (9.58, 9.97). Comparing the Predicted 2016 and 2020 Olympics' results that appear in the 9th and 10th columns in Table 1, most of the predicted results also fitted into the 95% P.I. Some of them did not do so, e.g. M100m Usain Bolt's both predicted results; W100m Shelly-Ann Fraser-Pryce's 2020 prediction result. Also some of the 95% P.I.'s appear to have very wide range between its Lower Bound and Upper Bound, as some parameter estimates are highly correlated.

The prediction models in Table 2 use the top 5 runs and hence present poor fits. We compared the predicted winning (PW) to actual 2012 Olympics Personal's (AOP) results in $Time_{seconds}$. This appears in the 3rd and 4th columns of Table 2; it presents both columns' results were fitting into the Predicted 95% Confidence Interval of Regression Line (95% P.I.) in the 5th column. However the 95% P.I. had sometimes an extremely wide range between the Lower and Upper Bound. For example, Yohan Blake's M200m's 95% P.I., its Lower Bound is -20.52 and Upper Bound is 61.15.

As known earlier the prediction can be dangerous, since almost predicted models present extremely bad predicted results. But for example, Sanya Richards-Ross's W200m's predicted result can change our mind. The prediction model is following:

$$Time_{seconds}(Tournament) = 21.306 + 0.246 \cdot (Tournament) - 0.104(Tournament)^2 + 0.010 \cdot (Tournament)^3 \quad (7)$$

Then we model $Tournament = 13$ for predicting 2016 Olympics, and $Tournament = 14$ for predicting 2020 Olympics. But the outcomes were extremely bad, 22.35s and 22.75s.
In 6th and 7th columns of the table, they present the Minimum Simulation $Time_{seconds}$ and Simulation 95% Confidence Interval. Compare with the AOP and Simulation 95% C.I., we see the AOPs generally do not fit into the 95% C.I except for example, Yohan Blake's 200m, the AOP is 19.44s and the 95% C.I. is (19.26, 20.39).

Event	Athletes	Predicted Winning Time_seconds	Actual 2012 Olympics Personal Time_seconds	Predicted 95% C.I. of Regression Line	Minimum Simulation Time_seconds	Simulation 95% C.I.	Simulation Distribution	Predicted 2016 Olympics Time_seconds	Predicted 2020 Olympics Time_seconds	r^2 values
M100m	Usain Bolt	9.84	9.63	(9.73,9.94)	9.58	(9.58,9.97)	Beta General	10.23	10.51	0.910
M100m	Yohan Blake	9.87	9.75	(9.76,9.97)	9.73	(9.84,9.99)	Logistic	9.87	9.88	0.312
M100m	Justin Gatlin	9.89	9.79	(9.21,10.57)	9.73	(9.82,10.23)	Exact Value	9.79	9.66	0.032
M100m	Asafa Powell	9.76	11.99	(9.43,10.08)	9.60	(9.79,10.03)	Logistic	9.59	9.36	0.239
M200m	Usain Bolt	20.02	19.32	(18.60,21.45)	18.04	(19.131,20.414)	Logistic	20.20	20.46	0.530
M200m	Yohan Blake	19.50	19.44	(15.80,23.19)	19.26	(19.260,20.600)	Beta General	19.15	18.63	0.241
M200m	Wallace Spearmon	20.39	19.90	(19.21,22.57)	19.13	(19.766,20.431)	Logistic	20.69	21.09	0.292
M200m	Churandy Martina	20.16	20.00	(19.22,21.09)	18.95	(19.750,20.801)	Logistic	20.42	20.87	0.657
W100m	Shelly-Ann Fraser-Pryce	10.44	10.75	(9.68,11.20)	10.62	(10.70,11.28)	Triangle	10.04	9.49	0.467
W100m	Veronica Campbell-Brown	10.68	10.81	(10.33,11.03)	10.51	(10.74,11.05)	Logistic	10.46	10.15	0.453
W100m	Carmelita Jeter	10.88	10.78	(10.37,11.47)	10.67	(10.77,11.27)	Exact Value	10.71	10.47	0.319
W200m	Allyson Felix	21.32	21.88	(20.64,22.00)	20.85	(21.81,22.76)	Logistic	20.74	20.00	0.836
W200m	Veronica Campbell-Brown	23.16	22.38	(21.89,24.30)	21.52	(21.78,23.11)	Exact Value	24.10	25.41	0.465
W200m	Sanya Richards-Ross	22.13	22.39	(21.25,23.00)	21.93	(22.09,22.94)	Exact Value	22.10	22.13	0.518

Table 1: Men's & Women's 100m & 200m predicted winning time with particular athletes' last 10 runs and their own actual winning running time in 2012 Olympics

Table 3 presents the time models using group based data. Once again, the results provide us with variable results. The M1500m is simulated (min) to virtually the same time.

Event	Athletes	Predicted Winning Time_seconds	Actual 2012 Olympics Personal Time_seconds	Predicted 95% C.I. of Regression Line	Minimum Simulation Time_seconds	Simulation 95% C.I.	Simulation Distribution	Predicted 2016 Olympics Time_seconds	Predicted 2020 Olympics Time_seconds	r^2 values
M100m	Usain Bolt	10.86	9.63	(1.87,19.68)	9.79	(9.79,10.26)	Beta General	12.76	16.14	0.822
M100m	Yohan Blake	10.00	9.75	(7.71,12.29)	9.75	(9.75,9.97)	Exponential	10.42	11.11	0.935
M100m	Justin Gatlin	10.75	9.79	(4.39,17.16)	9.85	(9.85,10.26)	Shift	12.22	14.69	0.752
M100m	Asafa Powell	9.53	11.99	(3.74,15.33)	9.85	(9.85,10.02)	Beta General	8.95	8.02	0.528
M200m	Usain Bolt	19.25	19.32	(-18.11,56.61)	19.07	(19.36,20.88)	Exact Value	17.70	14.92	0.998
M200m	Yohan Blake	20.73	19.44	(-20.52,61.15)	19.26	(19.26,20.39)	Beta General	22.33	24.85	0.951
M200m	Wallace Spearmon	19.01	19.90	(16.50,21.53)	19.57	(19.76,20.54)	Exact Value	16.51	12.10	0.565
M200m	Churandy Martina	21.28	20.00	(7.64,34.93)	19.94	(19.94,20.49)	Beta General	23.85	28.25	0.828
W100m	Shelly-Ann Fraser-Pryce	10.22	10.75	(7.40,13.04)	10.25	(10.70,11.21)	Logistic	9.46	8.35	0.975
W100m	Veronica Campbell-Brown	10.69	10.81	(4.82,16.56)	10.80	(10.81,11.12)	Shift	10.43	10.02	0.522
W100m	Carmelita Jeter	11.49	10.78	(9.58,13.40)	10.68	(10.91,11.16)	Logistic	12.51	14.27	0.953
W200m	Allyson Felix	20.62	21.88	(20.40,20.85)	21.00	(21.86,22.66)	Logistic	18.86	16.25	1.000
W200m	Veronica Campbell-Brown	23.98	22.38	(7.51,40.45)	21.38	(21.68,22.97)	Exact Value	26.95	31.95	0.817
W200m	Sanya Richards-Ross	22.19	22.39	(-0.53,44.90)	22.09	(22.09,22.63)	Beta General	22.35	22.72	0.328

Table 2: Men's & Women's 100m & 200m predicted winning time with particular athletes' last 5 runs and their own actual winning running time in 2012 Olympics

Table 4 uses the winner's times from each event as its sole fit. column except M100m's AOP. However the 95% P.I.'s had extremely wide range between the Lower and Upper Bound. But there are some exceptions, for example M200m's 95% P.I., (19.26,20.84).

Event	Predicted Winning Time_seconds	Actual 2012 Olympics Winning Time_seconds	Prediction 95% C.I. of Regression Line	Minimum Simulation Time_seconds	Simulation 95% C.I.	Simulation Distribution	Predicted 2016 Olympics Time_seconds	Predicted 2020 Olympics Time_seconds	r^2 values
M100m	9.83	9.63	(9.54,10.12)	9.56	(9.795,10.220)	Normal	10.83	13.28	0.716
M200m	19.37	19.32	(18.49,20.13)	19.87	(19.897,20.823)	Uniform	19.99	21.85	0.695
M400m	45.39	43.94	(44.45,46.32)	44.12	(44.495,46.523)	Exact Value	52.43	67.70	0.691
M800m	100.53	100.91	(97.51,103.56)	95.64	(101.46,108.61)	Logistic	88.06	60.42	0.403
M1500m	207.42	214.08	(202.45,212.39)	208.54	(208.70,233.30)	Exponential	187.88	141.76	0.545
M5000m	763.32	821.66	(741.27,785.36)	722.59	(762.30,802.70)	Logistic	669.86	421.96	0.870
M10000m	1600.36	1650.42	(1526.89,1673.83)	1621.33	(1621.41,1633.25)	Exponential	1498.71	1225.77	0.564
M42195m	7518	7681	(6452,8585)	N/A	N/A	N/A	6971	6028	0.208
W100m	10.54	10.75	(10.12,10.96)	10.88	(10.922,11.594)	Invgauss	9.57	7.52	0.539
W200m	22.52	21.88	(21.97,23.06)	21.08	(22.181,23.331)	Logistic	25.24	31.13	0.762
W400m	49.96	49.55	(48.26,51.65)	48.62	(49.43,52.78)	Invgauss	51.44	53.69	0.305
W800m	116.83	116.19	(111.35,122.31)	113.16	(116.98,121.89)	Logistic	123.32	139.27	0.148
W1500m	237.77	250.23	(229.72,245.83)	233.74	(236.07,249.87)	Triangle	225.55	192.14	0.529
W5000m	867.20	904.25	(829.83,904.57)	875.62	(876.30,919.50)	Triangle	882.53	911.71	0.366
W10000m	1845.26	1820.75	(1563.02,2127.50)	1824.39	(1824.39,1869.28)	Beta General	1804.50	1691.14	0.015
W42195m	9084	7587	(8268,9900)	N/A	N/A	N/A	10144	12189	0.456

Table 3: Men's & Women's predicting time with Top 5s' Results per year and actual winning running time in 2012 Olympics

In 5th and 6th columns of the table, they present the Minimum Simulation $Time_{seconds}$ and Simulation 95% Confidence Interval. Compared with the AOP and Simulation 95% C.I., we see the AOPs generally do not fit into the 95% C.I, except for example, W1500m, the AOP is 250.23s and the 95% C.I. is (238.05, 248.54).

Event	Prediction Winning Time_seconds	Actual 2012 Olympics Winning Time_seconds	Predicting 95% C.I. of Regression Line	Minimum Simulation Time_seconds	Simulation 95% C.I.	Simulation Distribution	Prediction 2016 Olympics Time_seconds	Prediction 2020 Olympics Time_seconds	r^2 values
M100m	9.89	9.63	(9.71,10.07)	9.95	(9.952,10.241)	Exponential	10.08	10.59	0.697
M200m	20.05	19.32	(19.26,20.84)	20.10	(20.10,20.62)	Beta General	21.69	25.67	0.515
M400m	45.91	43.94	(44.61,47.21)	44.53	(44.767,45.977)	Exact Value	52.43	66.69	0.583
M800m	102.63	100.91	(100.22,105.04)	103.31	(103.35,109.58)	Exponential	92.78	71.26	0.417
M1500m	212.05	214.08	(204.46,219.64)	N/A	N/A	N/A	201.08	176.83	0.259
M5000m	765.28	821.66	(726.33,804.23)	768.16	(768.16,792.71)	Beta General	644.96	362.54	0.501
M10000m	1619.37	1650.42	(1545.57,1693.17)	N/A	N/A	N/A	1587.00	1472.68	0.494
M42195m	7518	7681	(6452,8585)	N/A	N/A	N/A	6971	6028	0.208
W100m	10.84	10.75	(10.49,11.19)	10.98	(10.982,11.440)	Exponential	10.21	8.86	0.223
W200m	22.68	21.88	(22.37,23.00)	21.95	(22.495,23.012)	Logistic	23.85	26.43	0.369
W400m	50.60	49.55	(48.74,52.46)	49.88	(49.90,53.05)	Exponential	52.85	56.95	0.239
W800m	120.31	116.19	(116.65,123.96)	117.53	(118.21,121.36)	Exact Value	128.02	144.62	0.144
W1500m	243.15	250.23	(236.72,249.58)	238.05	(238.05,248.54)	Beta General	231.47	197.39	0.681
W5000m	877.17	904.25	(846.76,907.57)	N/A	N/A	N/A	822.44	704.65	0.176
W10000m	1845.26	1820.75	(1563.02,2127.50)	N/A	N/A	N/A	1804.50	1691.14	0.015
W42195m	9084	7587	(8268,9900)	N/A	N/A	N/A	10144	12189	0.456

Table 4: Men's & Women's prediction time with Winners' Results per year and actual winning running time in 2012 Olympics

5. CONCLUSION

The purpose of the paper was to compare the predictions between regression and simulation for eight track events (100m, 200m, 400m, 800m, 1500m, 5000m, 10000m and 42195m), for both men and women. From the results above, we can see: 1) the predicted results under regression are quite volatile for most events and 2) the predicted results under simulation present different distributions, both at an athlete level and group level. It appears that the 95% Confidence Interval provides us with a rather variable range that often is quite different to regression. Notably, the forward work is to compartmentalise runs to model sectional times of athletes.

References

Ballerini, R. (1987). "Records in the presence of a linear trend." Advanced Applied Probability 19: 801-828.

Ballerini, R., & Resnick, S. I. (1985). "Records from improving populations." Journal of Applied Probability 22: 487-502.

Chatterjee, S., & Chatterjee, S. (1982). "New Lamps for old: An exploratory analysis of running times in Olympic Games." Applied Statistics 31(1): 14-22.

Deakin, B. (1967). "Estimating bounds on athletic performance." The Mathematical Gazette 51: 100-103.

Kennelly, A. (1906). "An approximate law of fatigue in speeds of racing animals." Proceedings of the American Academy of Arts of Science 42(15): 275-331.

Liu, Y. (1993). Evaluation and Prediction of World Records and Ultimate Performance in Track and Field. Department of Human Kinetics. Vancouver, The University of British Columnbia. The Degree pf Master of Physical Education: 93.

Lucy, L. B. (1958). "Future Progress in the Mile." IOTA 1: 8-11.

Mognoni, P., Lafortuna, C., Russo, G., & Minetti, A. (1982). "An analysis of world records in three types of locomotion." European Journal of Applied Physiology 49: 287-299.

Ratkowsky, D. A. (1983). "Nonlinear Regression Modeling." Marcel Dekker Inc.

Schutz, R. W., Carr, R., & Halliwell, B. (1975). "The prediction of world track and field records." Unpublished manuscript, University of British Columbia.

DECISION MAKING UNDER RISK IN SPORTS, GAMBLING AND FINANCE

Sam Glasson[a, b]

[a] *National Australia Bank*
[b] *corresponding author: sam.glasson@nab.com.au*

Abstract

Decisions of risk versus reward are a common feature in sport, gambling and finance. Indeed without risk, each of these domains would become exceedingly boring. Each roulette bet would lose 2.7% per spin, the higher skilled team would always win, and the stock market would go up the same amount day after day. Yet each time we assess whether a risk is worth taking on, we must weigh up the potential rewards.
This paper examines the links between what on the surface appear to be quite different areas. For example, analysis of simplified gambling or sporting games can inform the correct strategy of a more complicated financial problem.
Drawing from diverse sources such as blackjack, horse racing, Australian football, financial instrument pricing, and high-frequency stock trading, I show how many of the mathematical techniques used in one domain transfer to another. A historical perspective is offered, highlighting examples of applications that cross the domain boundaries. Finally, a simplified framework for decision making under risky conditions is presented based on a risk versus reward tradeoff that is relevant across all three areas.

Keywords: gambling, finance, sport, risk, return, Kelly criterion, utility, professional gambling

TOWARDS INFORMED WAGERING IN SPORT

Brendan Poots[a, b]

[a] Founder & CEO, Priomha Capital Pty Ltd, Melbourne VIC 3000
[b] Corresponding author: bpoots@priomha.com

Abstract

Global sports betting markets turnover will grow to over $US500 billion by 2015. The consolidation of corporate bookmakers, the loosening of gaming laws in the USA and the continued growth of global sports betting exchanges such as Betfair (www.betfair.com) and Betdaq (www.betdaq.com) will continue to provide the necessary impetus for industry growth.

Concurrent to this growth has been the increased viewing and readily available information of sport through the widespread communication of it through TV, the Internet and other media. Additionally the recording and collation of sports-specific data has similarly increased.

The convergence of these factors has catalysed an opportunity for a new approach to wagering; an approach that is grounded in data and statistics. When this new approach to informed wagering is implemented with a clear strategy and discipline and executed correctly the results can be consistent and positive.

One such sport where informed wagering can be very effective is the English Premier League (EPL). The EPL is the most popular football league in the world. It is also the most popular global sport in terms of wagering turnover with $US50 billion invested per annum on the league.

This paper provides a high-level overview of how a professional Sports Hedge Fund has been able to use an informed approach to wagering to profit from the EPL and in doing so provide significant returns to its investors and stakeholders.

Keywords: Sports betting, EPL, Sports Hedge Funds

1. INTRODUCTION

Priomha Capital Pty Ltd ("Priomha Capital")

Priomha Capital is a boutique funds management firm whose investment universe centres on sports and events. It was created with a view to offer investors an alternative product to shares, bonds, property and other "traditional" investment vehicles and asset classes. Priomha Capital was founded in 2010 by a senior business executive who identified the opportunity to secure superior returns through the application of mainstream portfolio and investment management techniques to the nascent industry of sports and event investment. Through extensive research, due diligence, fundamental analysis and the use of technology, Priomha Capital has been able to develop a rigorous system grounded in statistical data, much the same way as has been achieved in more traditional financial markets, to produce superior returns.

The English Premier League ("EPL")

The EPL is the premier domestic football competition in the world. Not only is it the competition that is the most popular by way of broadcast rights and global appeal it is also the most wagered on sporting competition in the world with over $US50billion invested annually. Importantly, as the global appeal of the competition broadens so too will the interest from recreational and professional wagerers.

THE PATH TO INFORMED WAGERING

The growth in global wagering has catalysed the increase in readily accessible data pertaining to sport. This is particularly the case with the EPL where there are numerous dedicated websites and companies that collate, compile and present data in a format that is easily configured into a useable form. These explicit factors when combined with implicit factors that are ascertained to be relevant by the user provide the basis for an algorithm and mathematical model that help form the foundation of an approach to informed wagering.

Explicit Factors

The explicit factors pertaining to the EPL are omnipresent and are easily sourced. These include commonly known statistics relating to goals scored, goals conceded, home & away records and win/loss records.

Implicit Factors

The implicit factors used in modelling the EPL tend to be "user-specified" and as such are not limited in number. Typically however these would include the subjective weighting of past performances, the effects of weather on a potential outcome as well as idiosyncratic analysis of specific player match-ups, game strategy and individual match importance. Furthermore, and a critical factor is to consider the effects weight of money from the passion that is ubiquitous in the sports wagering markets has on probabilities and implied odds.

The Algorithm & Mathematical Model

Once the data is gathered a combination of COUNTIFS, SUMIFS and AVERAGEIFS functions are utilised as the basis of the model. Additionally, conditional probabilities, error checking and the use of iterative techniques to optimise the modelling of implicit factors are undertaken in order to complete the model.

2. RESULTS

The output of the algorithm, allows with some degree of confidence the ascertaining of probabilities of outcomes of certain events, be it for an entire season, match or any 25-minute interval within any of the 722 games that constitute a regular EPL season. Table 1 over summarises some of the obvious and more obscure statistics the analysis has elucidated whilst Figure 1 shows the performance of Priomha Capital's CLONEY Fund since inception.

Selected Event or Outcome to be Determined	***Top Ranked Team***	***Outcome****
What team wins the most matches?	Manchester United	67.20% (1.49)
What team is most likely to score first?	Manchester United	67.72% (1.48)
What team is most likely to win after scoring first	Manchester United	87.50% (1.14)
What team outperforms its average in wet conditions?	Newcastle United	+5.75%

* The number in parenthesis represents the equivalent decimal odds for the implied probability calculated

Table 1: Typical data generated by the algorithm

Priomha Capital's CLONEY Fund is a multi-sport investment fund. Its mandate is open to trade on any sport or event. Since inception in 2010 trading on the EPL has constituted over 32% of the total turnover of the Fund. Subsequently, whilst the graph shown in Figure 1 below is representative of the performance of the overall Fund, it can be viewed as a précis of the performance of trading in the EPL given the significant proportion of business that is done on that code.

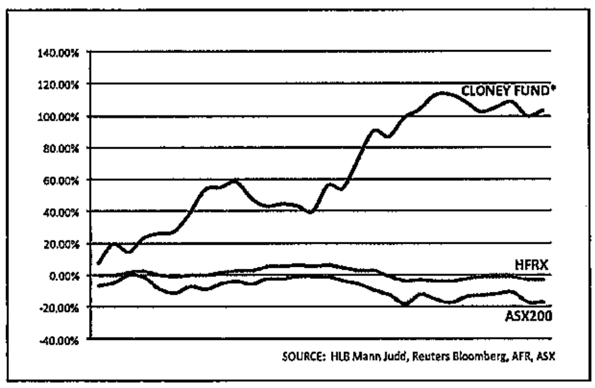


Figure 1: Performance of the CLONEY Multi-sport Fund

CONCLUSIONS & FURTHER WORK

The use of data, statistics and algorithms will never be able to be solely used as the means to undertake professional wagering successfully, as long term profitable wagering will always be a blend of art and science. The use of algorithms however is essential as the move towards informed wagering on sport continues to grow. The amount of readily available information warrants that a sophisticated means by which to analyse data is essential.

This is further strengthened by the fact that as corporate bookmakers and other service providers in the wagering industry become more advanced, one will need to continually revise, update and optimise databases and how they are used. The "perfect" algorithm will never exist but through continued updating and optimisation of databases and underlying assumptions an algorithm that continues to provide for a foundation that is profitable can be assured.

SPORTSMAN TRAINER AND AUTOMATIC SPORT- BRANCARD USING CLOSED LOOP CABLE SUSPENDED ROBOT

M. H. Korayem [a, b], H. Tourajizadeh [a]

[a] Robotic Research Laboratory, College of Mechanical Engineering, Iran University of Science and Technology, Tehran, Iran

[b] *Corresponding author: hkorayem@iust.ac.ir*

Abstract

A new mechanism is presented in this paper for simulating the athlete performance and training the sportsman's exercises, using a closed loop six degrees of freedom (DOFs) cable suspended robot. This robot cancels the necessity of presence of a sport coach for training the sportsman. Using the proposed robot, it is possible to program the robot for training the athlete limb (arm, leg and etc.) within a predefined trajectory corresponding to his special sport performance. The limb of the sportsman which is involved in the game and should be trained could be attached to the end-effector of the cable robot. Since in many sports, a large environmental space needs to be covered by the athlete movement, ordinary robots are not capable to be employed for this application while cable robots are applicable since a large dynamic workspace can be covered by them. Moreover training the sportsman limb requires a precise movement of the mentioned end-effector on a predefined trajectory. This importance could not be satisfied without using a proper closed loop controlling system since a variable external disturbing force applies on the end-effector as a result of the weight of the sportsman limb and its dynamic movement. Studio cams and automatic brancard for carrying the damaged sportsman out of the field are also of other applications of the presented closed loop cable robot. So required dynamic and control formulation of the end-effector of the cable robot is derived for handling the athlete limb on a predefined trajectory in a closed loop way. Simulation on the MATLAB confirms the possibility of the mentioned claim for simulating the sportsman training. Finally the efficiency of the proposed mechanism in training the athletes' limb is also proved by conducting experimental test on Iran University of Science and Technology (IUST) cable robot (ICaSbot).

Keywords: Sportsman training, Automatic sport-brancard, Studio cam, Cable suspended robot, Closed loop tracking control

References

Korayem M. H., Bamdad M., Tourajizadeh H., Shafiee H., Zehtab R. M., Iranpour A.,(2011), Development of ICaSbot a Cable Suspended Robot with 6 DOFs, *Paper accepted in Arabian Journal, 2011.*

Korayem M. H., Tourajizadeh H., Bamdad M., (2010), Dynamic Load Carrying Capacity of Flexible Cable Suspended Robot: Robust Feedback Linearization Control Approach, *Springer Journal of Intelligent and Robotic Systems.*

Roman, M. (1999). Flight Simulators: A Look at Linux in the Aerospace Training Industry. *Linux Journal, Vol. 19,* Available online at http://www.linuxjournal.com/.

Korayem M. H., Tourajizadeh H., (2011), Maximum DLCC of Spatial Cable Robot for a Predefined Trajectory within the Workspace Using Closed Loop Optimal Control Approach, *Journal of Intelligent and Robotic Systems.*

Zitzewitz, J. v., Rauter, G., Steiner, R., Brunschweiler, A., & Riener, R. (2009). A versatile wire robot concept as a haptic interface for sport simulation. *Proceeding IEEE International Conference on Robotics and Automation*, 313-318.

Korayem M. H., Jalali M., Tourajizade H., (2011), Optimal Path Planning of Spatial Cable-Suspended Robot Subject to Maximum Load Carrying Capacity Using Optimal Sliding Mode Control Approach, *3rd International Conference on Manufacturing Engineering ICME2011, Tehran, Iran*

AN APPLICATION OF BAYESIAN INFERENCE AND SIMULATION IN GOLF

Bradley O'Bree [a], Anthony Bedford [a,b], & Adrian J. Schembri [a]

[a] *School of Mathematical and Geospatial Sciences, RMIT University*
[b] *Corresponding author: anthony.bedford@rmit.edu.au*

Abstract

Professional golf has long been regarded as a difficult sport to model. The volatility of ever-changing conditions of play, such as weather, makes predicting a player's round score challenging. The relatively large playing field in any given tournament also makes the task of predicting each player's final position an ambitious one. A wealth of performance statistics correlated with success are available for elite tournaments in the United States. In this work, we propose a method for modelling professional golf tournaments using only historic round score data and round by round placings. Using the 2012 US Masters as a case study, we developed a methodology for predicting round scores for each player through possible scoring distributions. These scoring distributions were estimated using historic data from the competing players. A player's score was generated randomly from a distribution with the likelihood estimated from the player's observed historic score data. We simulated the tournament round by round, updating each likelihood using a Bayesian analysis of current score. Each simulation provides a measure of the probability of success for each player. These probabilities were seen to converge as the tournament was played out and each player's actual score became known. We validated the model's effectiveness by comparing the predicted outcomes with actual outcomes and those predicted from publicly released market prices.

Keywords: Golf, Bayesian, Simulation, US Masters

1. INTRODUCTION

Golf is a club and ball sport with worldwide popularity and origins dating back to at least the 15th century (Cherwoniak, 2008).

The objective is to complete each hole on the golf course in as few strokes (shots) as possible. Each course contains 18 holes, with each hole containing a tee off zone and a green, with the cup being located on the green. Professional players typically score an eagle, birdie, par, bogey or double bogey on each hole. Par refers to the number of strokes a professional player is expected to play to complete a hole, which is primarily judged on the length of the hole. Eagle and birdie refer to completing the hole in two or one strokes below par respectively. Bogey and double bogey refer to completing the hole in one or two strokes above par respectively.

Typically tournaments consist of four rounds, where playing 18 holes constitutes completing a round. After the second round, approximately half of the competing field of players is 'cut' from the tournament, meaning their participation is ceased. Competitors are ranked in an ascending order based on stroke counts, and those with the lower rankings (i.e. higher stroke counts, and therefore worst scores) are the ones who are cut from the tournament.

Typically performance modelling in golf is related to analysis of longitudinal data. Such performance statistics have been found to be highly correlated with scoring average and other measures important to success (Robertson, S, Burnett, A, Newton, R & Knight, P, 2012). Examples include driving accuracy and greens in regulation. Other variables, external to these performance statistics, such as competitive earnings, have been correlated with scoring average (Finley & Halsey, 2004). There is even evidence to suggest that there are home venue advantages in professional golf (Bailey, Clarke & Forbes, 2010).

The majority of work in golf prediction looks to model score (or scoring average) using these previously mentioned measures. The aim in this work was to develop a simulation model that accurately and efficiently predicts outcomes in professional golf tournaments without these measures. We sought to use only historic round scores and their associated results for players to simulate tournaments at the round by round level. Initially, this involved generating random round scores for each of the competing players at each stage of the tournament. In this work, we utilised the 2012 US Masters Tournament as a case study. This tournament is regarded as the pinnacle of not only the Professional Golfers' Association (PGA) Tour, but moreover all circuits worldwide. Only the best players in the world are invited to compete in this tournament.

2. METHODS

The simulation model followed the structure of the typical tournament (which is also valid for the US Masters). Scores were randomly generated for each player for each round in the tournament, with players ranked following each round (and the appropriate players cut following round two). Score distributions were dependent on each player's current score. The player with the lowest score at the end of round four was determined to be the winner.

2.1 Data

Data for the simulation model came in the form of round scores from completed professional tournaments. These were typically sourced from pgatour.com, the official website of the PGA Tour. This website provides scorecards and profiles for players competing on the PGA, Nationwide and Champions tours; all of which feature tournaments primarily in the US. Due to availability constraints, only scorecards dating back to the beginning of the 2011 season were acquired.

In consideration of specificity of any inferences from data used, we considered only the round scores of players competing in the tournament we were simulating. This means that while we collected scores from all currently competing players, when simulating the US Masters we only used the scores from players competing in the tournament.

Market prices were acquired from Bet365.com prior to the commencement of each round. These related to the likelihood a player would win the tournament outright. Prices are generally only released for those players with a chance of winning the tournament, meaning the sample size decreases as the tournament progresses. It should also be noted that these prices are subject to variation from public influence.

2.2 Software

All data manipulation and simulation was conducted using Microsoft Excel and Visual Basic for Applications (VBA) coding procedures, with some data analysis carried out using SPSS 18.

2.3 Round Score Distributions

We created a distribution of round scores using the round score data acquired for each of the competing players. A small transformation of the data was required, as round scores could relate to courses with differing course pars. We represent round scores as Par Percentage, which is equal to the ratio of the round score to its respective course par score. Course par scores very rarely deviate from being between 70 and 72, so taking the ratio shouldn't pose any problems, and is necessary to smooth the distribution (figure 1).

$$Par\ Percentage = \frac{Round\ Score}{Course\ Par\ Round\ Score} \quad (1)$$

These frequencies were standardised to create a probability distribution of round scores. Both these distributions can be approximated using a binomial distribution, the parameters of which are derived using the characteristics of the round score distribution.

Round scores for each round of the tournament can be randomly generated using this distribution. There is an issue however in that doing so assumes all players are of equal skill, which is obviously not the case. The objective then becomes to split the one scoring distribution into multiple distributions.

It was decided the simplest approach would be to have two different round score distributions for each round of the tournament. These distributions would be related to whether the player was expected to make the cut (not be cut after round two), and if they did so, the expectation they would finish in the top 10. Round scores were grouped according to the result for the player who scored them. In the former case, the first grouping contained scores where the player always made the cut. The second grouping contained scores where the player always failed to make the cut (figure 2). Having more than one scoring distribution allows for the quality of the player to be taken into account, as well as keeping the variation in scores for players reasonable.

Players would be randomly assigned to a scoring distribution, based on their inherent ability to achieve a score from that distribution. Using historic round score data, we determined the proportion of time the player's score belonged to each of the two distributions. These proportions were used as marginal probabilities to randomly assign a distribution to generate each player's round score.

2.4 Bayesian Inference

Bayesian inference was used to update these marginal probabilities while considering the player's current score in the context of the tournament. As we are generating round scores, at the end of each round the probability of assigning either distribution can be updated based on each player's current score. From here a new score could be generated for the following round.

The probability a score is sampled from either distribution is determined using Bayes' Rule (equation 2).

When simulating the first round, no observed data is available, so the probability either distribution is used is based purely on each player's marginal.

When simulating round two we have either observed or simulated round one scores, so we use these with the marginals to randomly assign a distribution to sample from. Here the distributions for the second round are split by whether the player made the cut. We use this same methodology for rounds three and four, however we split each round's distribution by whether the player placed in the final top 10, as the cut is made following completion of the second round. Choosing to split based on the eventually 10 best placed players was not optimised, it was an arbitrary number chosen to ensure the grouping was not too small.

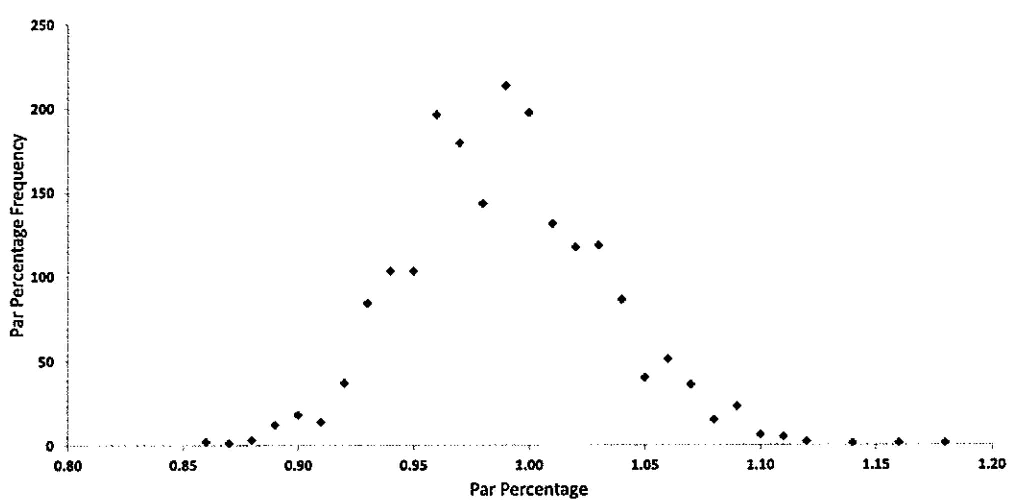


Figure 1. Distribution of acquired Round One Round Scores (Par Percentage) for competing players

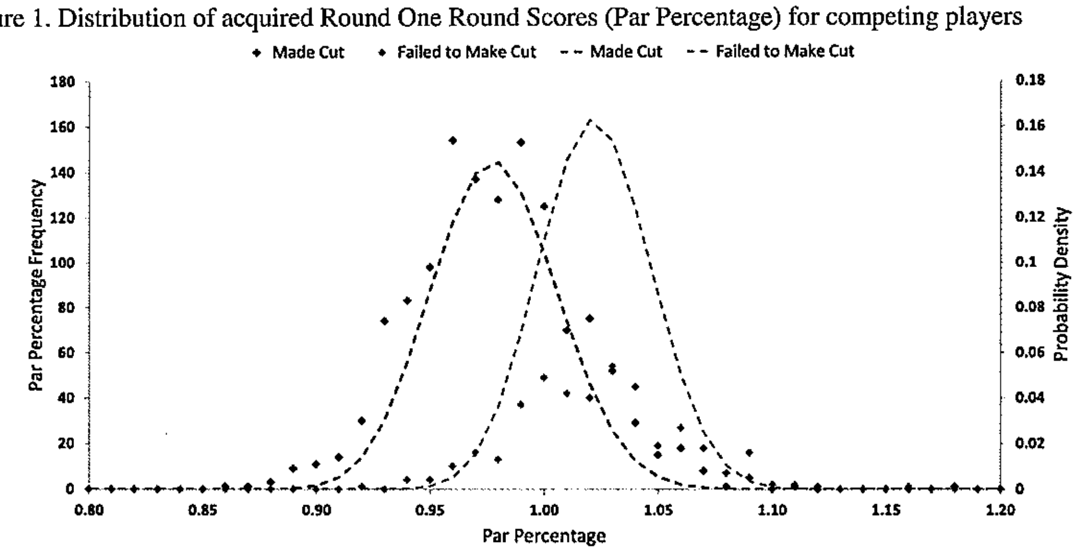


Figure 2. Distribution of acquired Round One Round Scores (Par Percentage) split by whether the player made the cut or not. Binomial approximations of the probability mass functions overlay.

$$P(\, Choose\ Distribution\ \theta \,|\, Current\ Score\ x) = \frac{P(x|\theta)P(\theta)}{P(x|\theta)P(\theta)+P(x|\theta')P(\theta')} \quad (2)$$

2.5 Simulation

In total we completed four simulations, each of 30,000 runs. The first was conducted pre-tournament and simulated all four rounds. The second simulated rounds two through four, using the observed round one scores in simulation. The third simulated rounds three and four the same way, and the fourth simulated just the final round.

In the event multiple players shared the lowest score at the conclusion of the fourth round, the winner was decided randomly taking into consideration their fourth round score only. This was done to capture the form of each player on the day. As we are generating scores at the round level and not the hole level, we cannot simulate any playoff holes, as would normally be the protocol after four rounds have been completed.

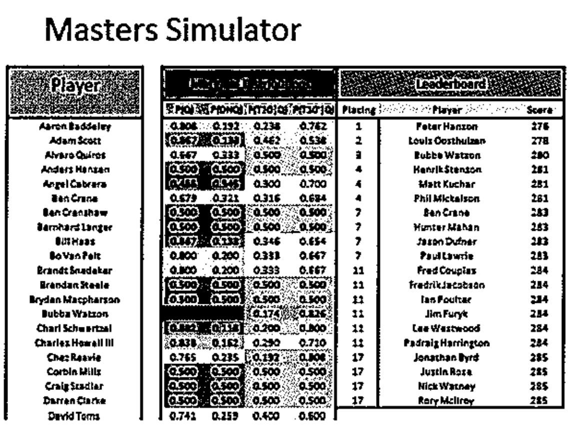


Figure 3. Screenshot of a simulation run

3. RESULTS

Given we are simulating scores for each player, a basic result to examine would be the appropriateness of inferred scores. It is better to look at the predicted ordering of players based on score, rather than the accuracy of absolute scores. We are generating scores at the round level and not the hole level, meaning it would be difficult to maintain a very good score prediction throughout the tournament, given the relatively large steps between estimations. This is especially the case when we look to simulate more than the last round of the tournament because errors would be combined for each additional round we simulate. We will examine the predicted final placings and the actual final placings for each simulation using the Spearman Correlation Coefficient. We will also examine the ability of this model to correctly place players as making the cut or placing in the top 10 using measures of classification and raw frequencies.

The proposed model, labelled RMIT, will be compared with two other models. Market prices for the outright winner were acquired from Bet365.com prior to each round of the 2012 Masters, and will be used to determine predicted final placings. While this conversion of prices is questionable, we are simply trying to gain an insight into the market that otherwise can't be realised due to restrictions on the data available. We do this because we can assume a relationship between winning odds and final placing. For example, the 10 players with the lowest market prices would be predicted to make up the top 10. The second is a naive model, which assumes that the current placings will be equivalent to the final placings. This model will provide the basis for comparisons of discriminative power.

The frequencies of correctly classified players, which regards to making the cut (table 1) indicate that when the tournament is simulated from its beginning the model is more effective than using the Bet365 market prices. This however was not the case once the observed first round scores were incorporated. Both models were only marginally more effective than the naive model.

Model	Players Correctly Assigned to Making the Cut#	
	Pre Round One	Pre Round Two
RMIT	51*	49
Bet365	44*	51*
Naive	-	48

* Jason Day was expected to make the cut but retired hurt during the second round

Total of 62 players made the cut

Table 1: Number of Players Correctly Predicted to Make the Cut

Model	Players Correctly Assigned to Placing in the Top 10#	
	Pre Round Three	Pre Round Four
RMIT	4	7
Bet365*	5	7
Naive	4	7

Total of 10 players were placed in the Top 10

* Bet365 Top 10 converted from outright win market lines

Table 2: Number of Players Correctly Predicted to Place in the Top 10

We observe a similar trend for classification of players placing in the top 10. The Bet365 market prices performed only slightly better than the proposed (RMIT) and naive models, classifying 5 and 4 of the top 10 players correctly respectively following the completion of round two. Interestingly, the three models classified seven of the eventual top 10 players correctly using the observed scores following round three. This suggests that the majority of the composition of the top 10 grouping was set by the end of the third round, and that the two models of comparison may not capture the information needed to identify the remaining group members. This could be the product of the elite playing field. Such a result and interpretation however would need to be verified using several tournaments.

We go further in measuring this discriminative power by calculating the specificity and sensitivity of these classifications between the models.

Classification for Players Making the Cut					
Measure	Pre Round One		Pre Round Two		
	RMIT	Bet365#	RMIT	Bet365#	Naive
Specificity	0.697	0.758	0.879	0.970	0.788
Sensitivity	0.645	0.694	0.742	0.823	0.758
Accuracy	0.663	0.716	0.789	0.874	0.768

Bet365 Top 10 converted from outright win market lines

Table 3: Measures of Players Correctly Predicted to Make the Cut following Round Two

Specificity is a measure of the model's ability to be correct when classifying a player as missing the cut. Sensitivity is a measure of the model's ability to be correct when classifying a player as making the cut. Accuracy is taken as the proportion of correct classifications.

With the focus on making the cut, we see the market prices were better at classifying players than the proposed model (table 3). This is seen in both relevant simulations, where simulation uses no observed round scores and where simulation uses only round one observed scores. Both models were seen to provide better classification than the naive model. An interesting note is that the specificity measure from the betting model using the round one observed scores was near perfect at 0.970, meaning it was virtually faultless in identifying the players who would miss the cut.

The trend that was seen when classifying players making the cut was seen when classifying the top 10. The market prices better classified players when compared to the proposed and naive models.

Classification for Players Placing in the Top 10						
Measure	Pre Round Three			Pre Round Four		
	RMIT	Bet365#	Naive	RMIT	Bet365#	Naive
Specificity	0.885	0.904	0.885	0.942	0.923	0.942
Sensitivity	0.400	0.500	0.400	0.700	0.500	0.700
Accuracy	0.806	0.839	0.806	0.903	0.887	0.903

Bet365 Top 10 converted from outright win market lines

Table 4: Number of players correctly predicted to Place in the Top 10

The measures of the proposed and naive models were identical for both relevant simulations. This suggests the proposed model may not yet provide any more explanatory power than is provided by the current placings when predicting the players who would finish in the top 10. The market prices however do.

Model	Final Placings Correlation Coefficient			
	Pre Round One	Pre Round Two	Pre Round Three	Pre Round Four
RMIT	0.135	0.274*	0.562**	0.708**
Naive	-	0.279*	0.547**	0.721**

* Correlation is significant at the 0.05 level

** Correlation is significant at the 0.01 level

Table 5: Spearman Correlation Coefficient for Rankings and Final Placings

We also examined the ability of the proposed model to correctly predict the placing of players at the end of the tournament. Here, we can only compare placings with the naive model due to restrictions on player field coverage for acquired market prices in the later rounds. The correlations between predicted final placings and actual final placings were determined for all four simulations (table 5). As one would expect, the later into the tournament, the stronger the correlations were found to be. They were essentially equivalent between the models, indicating again that the proposed model did not provide additional explanatory power. All correlations were statistically significant with the exception of the correlation from the first simulation.

4. DISCUSSION

Results from simulation of the 2012 Masters indicated that when comparing the proposed model with Bet365 market prices and a naive model, the proposed model had at least the explanatory power of the naive model, and in general less explanatory power than the Bet365 model. This was found both when predicting which players would make the cut and identifying which players would place in the top 10 at the end of the tournament.

We saw that when classifying players who would place top 10, the proposed and naive models were equivalent. It is not necessarily bad that the proposed model adds no explanatory power as it uses only historic round scores and the success rate for placing in the top 10. With the quality of the field of a tournament like the US Masters, we would not expect to see a great deal of movement in rankings between the third and final rounds. Those who are leading would be expected to remain composed, to the extent that those chasing are not expected to catch them. It may be in a tournament of lesser significance that there is more variation in the later stages of the tournament, hence the apparent lack of added explanatory power in the proposed model.

The final placings as predicted by the proposed model had a correlation of equivalent strength and significance with the observed final placings as that with the current observed placings for each simulation. These correlations became relatively strong as more observed scores were used.

The proposed model was not able to provide an improved insight into variation in standings throughout the tournament. The results however suggest it is a suitable base model for simulating golf tournaments. With further development and integration of explanatory variables, such as those performance measures that currently dominate performance modelling methods, predictions would surely improve.

A simple way the proposed model can be developed further is by shortening the steps between estimation. Simulating scores hole by hole would introduce more complexity into the overall simulation process, but at the same time would allow more accurate score prediction.

5. CONCLUSIONS

Results indicated that the proposed model was at least as effective as the naive model in predicting outcomes in the 2012 US Masters. At different times for different predictions, it was seen to be comparable with a model that used market prices released by Bet365. These results suggest the proposed model has a solid foundation for future development and improvement, something very promising for a model that currently samples from a data pool of such small dimensions.

References

Bailey, M, Clarke, S & Forbes, D 2010, 'Home Continent Advantage on the American and European professional Golf Tours', Proceedings of the tenth Australasian Conference on Mathematics and Computers in Sport, A Bedford & M Ovens (Eds), pp.173-180.

Cherwoniak, G. 2008, The History of Golf – Playing through the Centuries, www.thehistoryof.net.

Finley, P. S. & Halsey, J. J. 2004, 'Determinants of PGA tour success; An examination of relationships among performance, scoring and earnings', Perceptual and Motor Skills, 98, 1100-1106.

Robertson, S, Burnett, A, Newton, R & Knight, P 2012, 'Development of the Nine-Ball Skills Test to discriminate elite and high-level amateur golfers', Journal of Sports Sciences, 30:5, 431-437.

Shepherd, R, Bedford, A & Schembri, A 2010, 'A Two-Stage Simulation to Predict Medallists in Pistol Shooting', Proceedings of the tenth Australasian Conference on Mathematics and Computers in Sport, A Bedford & M Ovens (Eds), pp.213-220.

TALENT POOL MEMBERSHIP IN SPORT ORGANIZATIONS BY FUZZY ANALYTIC HIERARCHY PROCESS

Reza Kiani Mavi [a, c], Neda Kiani Mavi [b]

[a] *Assistant Professor, Department of industrial Management, Faculty of Management and Accounting, Qazvin branch, Islamic Azad University (IAU), Qazvin, Iran*
[b] *Department of Physical Education and Sport Science, Faculty of Management and Accounting, Qazvin branch, Islamic Azad University (IAU), Qazvin, Iran.*

[c] *Corresponding author: Rezakianimavi@yahoo.com*

Abstract

The ability to attract and retain talent is rapidly becoming one of the key issues for human resource managers and their organizations across the globe. For organizations across the globe, talent management of knowledge workers and managers is of strategic importance. Sport organizations are not exception from the case. This paper presents a fuzzy analytic hierarchy process (FAHP) for scouting and evaluation of young sport talents. Fuzzy logic is implemented in order to make the results more acceptable. Based on the knowledge of several sport experts, various attributes for membership in talent pool are ranked. Four main criteria in this paper are knowledge and experience, drive and energy, pride and responsibility and ability to deliver results. Findings show that knowledge and experience is the most important element for talent pool membership. Next, a linear function developed that can illustrate the suitability of candidates for talent pool membership.

Keywords: Talent management, Talent pool, Fuzzy AHP, Ranking

IN-GAME SIMULATION OF SCORE PREDICTION IN NBA

Jong Ho Park[a, b], Anthony Bedford[a]

[a] *School of Mathematical & Geospatial Sciences, RMIT University, Melbourne, VIC, Australia*
[b] *Corresponding author: s3261079@student.rmit.edu.au*

Abstract

The purpose of this study is to predict the outcome of a basketball game or after the quarter end result. The play-by-play data offered by the media was used for real-time basketball score and outcome predictions. The data used in this study were gathered from the 2009/10 NBA regular season, and we applied score and outcome prediction of games to the 2010/11 regular season. We predict the final results based on the various real-time basketball statistics extracted by the play-by-play data statistics. The playing time of basketball is 48 minutes in the NBA, so we divided a game into 8 time units, and we examine the score distribution in each section. The probability distributions are made from the home and away team score discrepancy at each time point and hence help us to determine the final results. The final and quarter scores are predicted using a smoothed probability distribution for each interval score distribution. The negative binomial functions of the interval score distributions were obtained according to the score discrepancy at each time point and home/away outcome, this score and outcome score prediction results were simulated. We tested its efficiency by comparing the betting market and will also confirm whether it will make profit or not in in-game prediction market in coming NBA season.
Keywords: In game, NBA, play-by-play data, probability distribution

1. INTRODUCTION

With the advent of in game data via the internet, sports statisticians have the power to test their models dynamically. Furthermore, with the ability to record data live from bookmakers and betting exchanges, an unprecedented opportunity is available to test models against public prices. A body of research exists on basketball models.

Much research in basketball prediction has been undertaken on NCAA basketball. Boulier et Stekler (1999), Caudill (2003), Caudill and Godwin (2002), Harville (2003), and Kaplan and Garstka (2001) used a seeding system as a predictor of outcomes in NCAA basketball tournaments. All prediction rates of these methods, which used the seeding difference, were around 70%. In particular, Boulier et Stekler and Caudill applied the probit model and the maximum score estimator individually as predictors in their model. However, their profit difference between the two models was not significant though the Caudill's had a slight better prediction rates. On the other hand, Zak, Huang, and Siegfried (1979) built a predictive model that represents the defensive and offensive elements of teams. Their model explained the final rank of all NBA teams based on the winning probability very well. Berri (1999) calculated the contributions of each player, and used the sum of their contributions as an estimator of the team rank. Carlin (1996) estimated the point difference of basketball games using seeding numbers and Sagarin rating differences in NCAA tournament. They compared the actual point difference with the seeding or rating difference, obtain regression equations in each case. Strumbelj and Vracar (2012) stated that the increasing accessibility of in-play betting had brought a demand for real-time forecast based on the progress of the sporting event. In addition to the simple outcome of matches, they are interested in the specific records such as score, rebound in basketball game.

Recently, in-play modelling has emerged as live match data is ever-present with the development of information technology. In this research, we suggest a simulation approach to the final score or a score at the mid and end of each quarter after the game already starts. This 'real time' score prediction is the one of exciting and interesting subjects in sports prediction. In this research, we investigate the nature of in-game scoring in NBA basketball. Through analysis of data from 2009/10, we look at the nature of scoring patterns. Applying modelling methods in Shepherd and Bedford (2010), we look to model and simulate match scores in-play.

2. METHODS

A number of aspects to modelling NBA scores are important. Firstly, the nature of data needs to be determined. We decided to simulate our model derived from the probability distribution of all past scores of fixed time interval because we judged that this kind of probability model was more appropriate for a large number of score game such as basketball. This research uses only score information in specific time in 2009/10 NBA regular season 1226 match data. We collected score data in each 6 minutes unit. Only each 4 quarter score are recorded in a basketball score board. So, we found the extra score information per each half quarter in the play-by-play transaction data. Table 1 shows the raw data for our score prediction in detail.

H/A Team	Home (Cleveland)	Away Boston
End of 1HQ (6 minutes)	19	7
Score margin	12	-12
Added Point between 1HQ and 1Q	9	14
End of 1Q (12 minutes)	28	21
Score margin	7	-7
Added Point between 1Q and 2HQ	8	13
End of 2HQ (18 minutes)	36	34
Score margin	2	-2
Added Point between 2HQ and 2Q	9	17
End of 2Q (24 minutes)	45	51
Score margin	-6	6
Added Point between 2Q and 3HQ	7	13
End of 3HQ (32 minutes)	52	64
Score margin	-12	12
Added Point between 3HQ and 3Q	13	8
End of 3Q (36 minutes)	65	72
Score margin	-7	7
Added Point between 3Q and 4HQ	10	8
End of 4HQ (42 minutes)	75	80
Score margin	-5	5
Added Point between 4HQ and 4Q	14	15
End of 4Q (48 minutes)	89	95
Score margin	-6	6

Table1. Data used in every half quarter. (Cleveland vs Boston 10/27/2009)

2.1. Basic model

The in-game model was split into eighths so that we effectively model the game earlier than at the end of the first period. So, our model simulates after completion of 1/8 of play (6 minutes in), for each eighth, until the 7/8 of play (6 minute remaining in match). It is interesting to note that many in-game markets will close at different times, especially as either the match is nearing its end, or when results are beyond reasonable doubt (i.e. large margins).

Let us denote the predicted score PS_i and PS_j of home and away team i,j at specific time k. The fixed score $FS_{i,k}$ and $FS_{j,k}$ at a specific time k indicates the actual scores of home and away team i,j after the game already starts. Let $X_{i,k,k+1}$ and $X_{j,k,k+1}$ at specific time k be the simulated unit interval score. These score probability functions are generated from the collected score data between unit time intervals in 1226 matches. The k value indicates the specific time of 1(6min.), 2(12min.), 3(18min.), 4(24min.), 5(30min.), 6(36min.), 7(42min.), and 8(48min.). We can predict each time point of score from 12 min. to the final (48min.) result. The following model is the predicted score of home and away team at time k+1.

$$PS_{i,k+1} = FS_{i,k} + X_{i,k,k+1} \quad \text{(Home)} \quad (1)$$

$$PS_{j,k+1} = FS_{j,k} + X_{j,k,k+1} \quad \text{(Away)} \quad (2)$$

where $0 \leq i \leq 30$, $0 \leq j \leq 30$, $1 \leq k \leq 7$

The final score model is the sum of the fixed score and the predicted one from time k+1 to the final time.

$$PS_{i,final} = FS_{i,k} + \sum_{l=k}^{7} X_{i,l,l+1} \quad \text{(Home)} \quad (3)$$

$$PS_{j,final} = FS_{j,k} + \sum_{l=k}^{7} X_{j,l,l+1} \quad \text{(Away)} \quad (4)$$

where $0 \leq i \leq 30$, $0 \leq j \leq 30$, $1 \leq k \leq 7$

Given that the data was split into eighths, we needed to determine the underlying distribution of scores. However, given we were utilising conditional distributions, we needed to fit data distributions to historical data for each 1/8 of each match, split by home and away. Ideally, we would model this for each team, however this initial work looks solely at the teams as a set. The model used is a margin based approach. A number of considerations were made. Firstly, if we use a score based approach, then prediction are essentially independent of the conditions of the game. For example, if we model a home side scores for the time 1/8 to 2/8, this does not consider the state of the game, and may well over-estimate a losing teams ability to score based on their current predicament. Margin is a (dependent) measure of the game condition, and provides our model with a context of the game, rather than simply the scoring of a team.

2.2. Winning probability distribution

Let us define the winning probability distribution as follows.
Let

$$\Omega_i = P(x|k = q, d = m) \qquad (5)$$

where x is the match outcome, k is the specific time from a half quarter to a forth and half quarter, and d is the score margin.
Ω denotes the winning probability of a team at specific time k at margin d. This set of probabilities will be used in the simulation for score prediction. Table 2 shows the winning probability table at k=1.
We investigated the relationship between the margin at the time 1/8 and the winning probability of each team. At a margin of -10, the home team at final records 8 wins, 4 draws, and 10 losses (seen in table 2). Excluding drawn matches, the winning probability is 0.444. On the other hand, at a margin of +10, the winning probability of a home team at final is 0.775.

Margin	W	D	L	WP	LP
.					
.					
-10	8	4	10	0.444	0.556
-9	11	0	13	0.458	0.542
-8	7	3	15	0.318	0.682
-7	12	5	20	0.375	0.625
-6	21	2	14	0.600	0.400
-5	15	4	24	0.385	0.615
-4	18	2	28	0.391	0.609
-3	26	3	25	0.510	0.490
-2	32	5	33	0.492	0.508
-1	38	4	22	0.633	0.367
0	37	6	38	0.493	0.507
1	32	3	20	0.615	0.385
2	41	1	31	0.569	0.431
3	42	4	21	0.667	0.333
4	41	1	13	0.759	0.241
5	47	5	11	0.810	0.190
6	32	3	11	0.744	0.256
7	29	3	11	0.725	0.275
8	28	1	4	0.875	0.125
9	20	2	8	0.714	0.286
10	31	0	9	0.775	0.225
.					
.					

Table2. The margin and winning probabilities at the time 1/8

It is obvious that the winning probabilities of home team are increasing as the score differences are also increasing. At a score difference of +10 after the end of 3rd quarter of table 3, its winning probability of a home team is 0.920, which is higher than that of the end of 1st quarter. At a score difference of -10, its value is only just under 0.100. As the game approach its conclusion, it shows that the outcome is more decisive than in earlier time, as expected.

Margin	W	D	L	WP	LP
.					
.					
-10	2	1	22	0.083	0.917
-9	1	1	19	0.050	0.950
-8	8	2	18	0.308	0.692
-7	6	0	29	0.171	0.829
-6	8	3	14	0.364	0.636
-5	4	4	19	0.174	0.826
-4	17	6	28	0.378	0.622
-3	17	5	18	0.486	0.514
-2	13	6	18	0.419	0.581
-1	21	2	16	0.568	0.432
0	21	3	13	0.618	0.382
1	13	9	20	0.394	0.606
2	22	2	22	0.500	0.500
3	22	3	8	0.733	0.267
4	26	3	9	0.743	0.257
5	34	5	8	0.810	0.190
6	26	5	5	0.839	0.161
7	28	2	3	0.903	0.097
8	28	2	1	0.966	0.034
9	34	0	1	0.971	0.029
10	23	1	2	0.920	0.080
.					
.					

Table3. The margin and winning probabilities at the time 6/8

2.3. Data fits

The winning probabilities at time k are a little bit noisy because they did not show a consistent profile. This is due to the sparsity of outcomes on the fringes, or in some instances, a simple lack of outcomes at all. Sargent and Bedford (2009) applied a non-linear smoother to AFL player data for removing noise, and applying some sense around values between sparse values. Here a Tukey T4253H smoother was imposed on the winning probability distribution (an example is seen in table 4).
SWP are the smoothed winning probabilities and SLP the smoothed losing probability. We exclude the draw term when we calculate the winning and losing probabilities for simplicity in the simulation algorithm.

Margin	W	D	L	WP	LP	SWP	SLP
.							
.							
-10	9	0	31	0.225	0.775	0.238	0.763
-9	8	2	20	0.286	0.714	0.245	0.755
-8	4	1	28	0.125	0.875	0.236	0.764
-7	11	3	29	0.275	0.725	0.262	0.738
-6	11	3	32	0.256	0.744	0.268	0.732
-5	11	5	47	0.190	0.810	0.323	0.677
-4	13	1	41	0.241	0.759	0.383	0.617
-3	21	4	42	0.333	0.667	0.408	0.592
-2	31	1	41	0.431	0.569	0.427	0.573
-1	20	3	32	0.385	0.615	0.472	0.528
0	38	6	37	0.507	0.493	0.503	0.497
1	22	4	38	0.367	0.633	0.529	0.471
2	33	5	32	0.508	0.492	0.554	0.446
3	25	3	26	0.490	0.510	0.581	0.419
4	28	2	18	0.609	0.391	0.610	0.390
5	24	4	15	0.615	0.385	0.602	0.398
6	14	2	21	0.400	0.600	0.587	0.413
7	20	5	12	0.625	0.375	0.608	0.392
8	15	3	7	0.682	0.318	0.650	0.350
9	13	0	11	0.542	0.458	0.640	0.360
10	10	4	8	0.556	0.444	0.688	0.312
.							
.							

Table4. Smoothed winning probability at the time 2/8

In fact, the unit score probability distributions for summation at fixed score at specific time k does not have a complete probability function form. So, we utilized the @Risk package to fit score distributions. We obtain the fitted four probability distributions in each unit time division.

This winning probability is the standard value for selection of probability distribution functions. We simulate the random number between 0 and 1. And then this arbitrary number is compared to the winning probability of the margin at time k. If this randomly generated number is over the winning probability value, the outcome may be "Home lost and Away won", otherwise it will be "Home won and Away lost". The corresponding probability distribution functions in home and away teams are determined by the outcome at specific time k. The generated random number chooses the outcome at time k and the corresponding unit score negative binomial probability functions between time k and k+1 after it is compared to the winning probability at time k. The unit score negative binomial probability functions are $NB_{i,k,k+1,HW}(\cdot)$, $NB_{i,k,k+1,HL}(\cdot)$, $NB_{j,k,k+1,AL}(\cdot)$, and $NB_{j,k,k+1,AW}(\cdot)$ between time k and time k+1 □ Home won and Away lost, Home lost and Away won. The predicted unit interval score from time k to k+1 is categorized to two cases. Therefore, the predicted score $PS_{i,k+1}$ and $PS_{j,k+1}$ at time k+1 will be the sum of the fixed score at time k and average values of those functions for each home time i, away team j.

$$\begin{cases} X_{i,k,k+1,HW} \sim NB_{i,k,k+1,HW}(n_{i,k,k+1}, p_{i,k,k+1}) \\ X_{j,k,k+1,AL} \sim NB_{j,k,k+1,AL}(n_{j,k,k+1}, p_{j,k,k+1}) \end{cases}$$
(Home Won & Away Lost) (6)

$$\begin{cases} X_{i,k,k+1,HL} \sim NB_{i,k,k+1,HL}(n_{i,k,k+1}, p_{i,k,k+1}) \\ X_{j,k,k+1,AW} \sim NB_{j,k,k+1,AW}(n_{j,k,k+1}, p_{j,k,k+1}) \end{cases}$$
(Home Lost & Away Won) (7)

2.5. Modelling process

The basis of our modelling is as follows:

1. Determine the underlying nature of scoring patterns for all teams in each 1/8th for home and away
2. Ascertain the likelihood of winning for home (away) at each 1/8th based upon the margin of the game
3. Smooth likelihoods
4. Establish scoring distributions for teams based upon these margins from historical data
5. Smooth fits
6. Simulate matches in-play based upon existing margin information for each stage
7. Evaluate the errors at each stage.

2.6. The Obvious Problem with Score Modelling

A number of fits to score were conducted first. Notably, and somewhat obviously, clustering scoring data and determining empirical likelihoods would never work-but we wished to see if there was at least some underlying interaction.

3. Results

All probability function distributions fit were from the NBA 2009/10 season's data. So, we tested our model in future games, and simulated 10,000 times within each match. We usually obtain seven results in a match in each simulation. The first half of the first quarter score is required for simulation, as we have no pre-game estimates at this stage of development.

At first, we fixed a first-half quarter (six minutes after a game starts) score and sum the other six simulated scores to collect the final score. As the game progresses, we fixed a first quarter score (12 minutes after a games starts) and sum other five simulated scores, and so on.

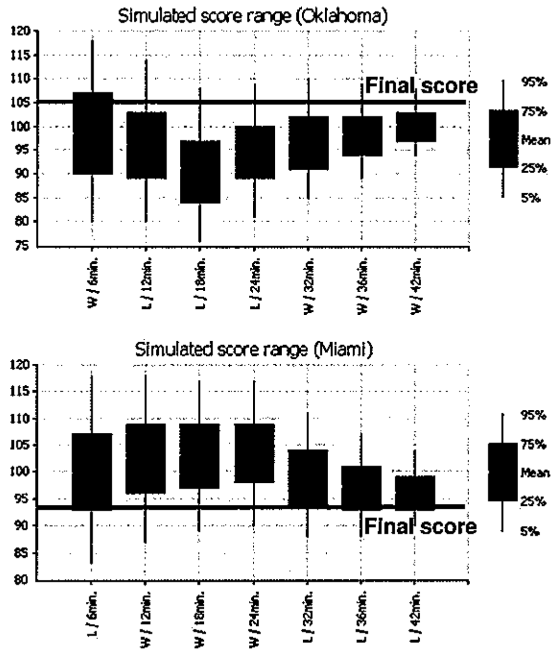


Figure 1. The Box plots of final score of Oklahoma City and Miami in every half quarter. (Oklahoma vs Miami 6/12/2012)

Figure 1 shows the seven score results of Oklahoma City Thunder and Miami Heat in the 1st game of NBA final series match of 2011/12 season. Oklahoma City turned the game around in the 3rd quarter, with its score at 105, and Miami's 94. As playing time goes on, the score error ranges are naturally smaller, and the simulated mean scores approach the actual scores closer.

Time	Oklahoma Error	Miami Error
Fixed from 1HQ (6 minutes)	-6.4	+5.9
Fixed from 1Q (12 minutes)	-8.8	+8.1
Fixed from 2HQ (18 minutes)	-13.8	+9.1
Fixed from 2Q (24 minutes)	-10.4	+9.4
Fixed from 3HQ (32 minutes)	-8.3	+5.1
Fixed from 3Q (36 minutes)	-6.4	-3.3
Fixed from 4HQ (42 minutes)	-4.7	+2.1
End of 4Q (48 minutes)	0.0	0.0

Table 5. Score error between the actual and simulated in 1st match of 2011/12 season NBA final.

We also investigated the score error profiles between the actual score and the simulated score in Table 5. The simulated scores are the summation of mean values from the score probability functions of each unit time. Its score error variation is dependent on the flow of that day. The errors of this game are maximized at the second quarter, and the variation approaches to zero after Oklahoma dominated the game in the third quarter. Generally, the errors are smaller as a game is approaches its conclusion. But, there are a few exceptions where the errors increase near game's end. We found the property of the score distributions of both teams as time passed and illustrate this in Figure 2. The output reveals the mean scores of Oklahoma City, 98.6 (6 min.), 94.6 (18 min.), and 100.3 (42 min.).

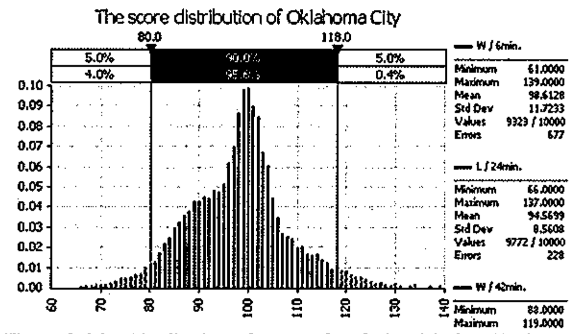


Figure 2. The Distribution of score after 6min., 24min., 42min.

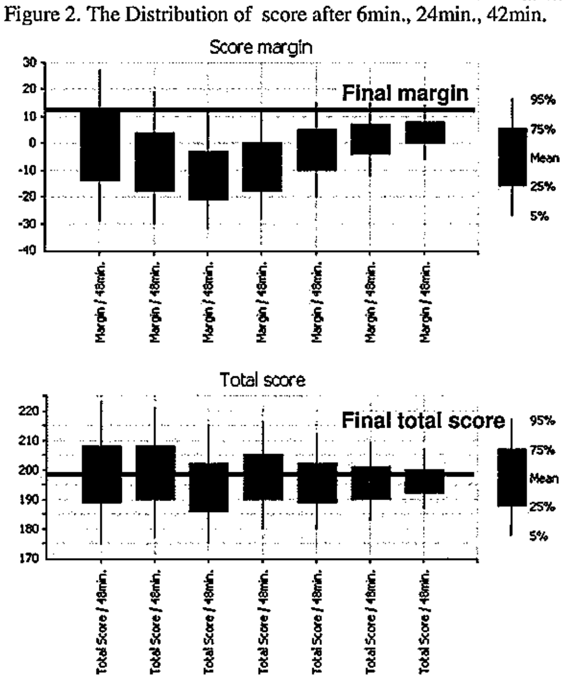


Figure 3. The Box plots of margin and total score. (Oklahoma vs Miami 6/12/2012)

Figure 3 is a box plot graph of the score margin and total score between two teams. Its margin profile is approaching the actual margin as the game comes to an end. On the other hand, total score is almost constant regardless of its playing time progress in spite of fluctuations of each score. But, a predicted total score or margin in play is not in this instance close to the actual score. The reason why the total score does not change a lot above match is that the home team score has a tendency to increase and away team score has a tendency to decrease.

4. DISCUSSION

Although this score simulation method only use scores, our prediction results are quite close to real scores in our results. However, there are a few things required to improve our model. We do not include many factors such as shots, rebounds, and ball possession statistics. The next study is to investigate how these factors other than score have an impact on the in-game score prediction and how efficient we can integrate probability functions. The variation or tendency of these factors will give us more important clues for more exact prediction.

Another important aspect that is not included is the tempo of the game, which is the probable reason for the lack of quick convergence in the some of our simulations.

We did not test its efficiency in basketball betting markets. In real time betting, it is well known that the total score and score margin market is one of the biggest basketball betting markets. We will also plan to estimates its usefulness in 2012/13 season.

5. CONCLUSIONS

We tested a real time score prediction model in basketball match using stage based negative binomial functional fits. We found that this method is quite suitable for predicting exact score, and has potential for other aspects of basketball modelling, including short-term estimates, margin and total score. With further back testing, and forward, we hope to be able to tighten these results and yield a system that produces significantly more accurate results of NBA outcomes.

References

Berri, D. J. (1999). Who is 'most valuable'? Measuring the player's production of wins in the National Basketball Association. *Managerial and Decision Economics,* 20, 411-427

Boulier, B. L., Stekler, H. O. (1999). Are sports seedings good predictors? : an evaluation. *International Journal of Forecasting,* 15, 83-91

Carlin, B. P. (1996). Improved NCAA basketball tournament modelling via point spread and team strength information. *The American Statistician, 50*, 39-43

Caudill, S. B. (2003). Predicting discrete outcomes with the maximum score estimator: the case of the NCAA men's Basketball tournament. *International Journal of Forecasting,* 19, 313-317

Caudill, S. B., Godwin, N. H. (2002). Heterogeneous skewness in binary choice models: Predicting outcomes in the men's NCAA basketball tournament. *Journal of Applied Statistics*, 29, 991-1001

Harville, D. (2003). The selection or seeding of college basketball of football teams for post season competition. *Journal of the American Statistical Association*, 98, 17-27

Kaplan, E. H., Garstka, S. J. (2001). March madness and the office pool. *Management Science*, 47, 369-382

Sargent, J., Bedford, A. (2010). Improving Australian Football league player performance forecasts using optimized nonlinear smoothing, *International Journal of Forecasting,* 26, 489-497

Shepherd, R., Bedford, A., Schembri, A. (2010). A two-stage simulation to predict medallists in pistol shooting. *Proceedings of the Tenth Australasian Conference on Mathematics and Computers in Sport (10 M&CS),* Australia, 5-7 July 2010, 213-220

Strumbelj, E., Vracar, P. (2012). Simulating a basketball match with a homogeneous Markov model and forecasting the outcome. *International Journal of Forecasting,* 28, 532-542

Zak, T. A., Huang, C. J., Siegfried, J. J., (1964). Production efficiency : The case of Basketball. *Journal of Business*, Vol. 52, No. 3, 379-392

THE EFFECT OF RISKS ON SPORT ORGANIZATIONS USING DEMATEL

Arezoo Jalili [a,*] , Somayeh farid [b] , Reza Kiani Mavi [c], Neda Kiani Mavi [d]

[a]*Department of Industrial Management, Faculty of Management and Accounting, Qazvin branch, Islamic Azad University (IAU), Qazvin, Iran.*
[b]*Department of Industrial Management, Faculty of Management and Accounting, Qazvin branch, Islamic Azad University (IAU), Qazvin, Iran.*
[c]*Assistant Professor, Department of Industrial Management, Faculty of Management and Accounting, Qazvin branch, Islamic Azad University (IAU), Qazvin, Iran.*
[d] *Department of Physical Education and Sport Science, Faculty of Management and Accounting, Qazvin branch, Islamic Azad University (IAU), Qazvin, Iran.*

[*]*Corresponding author: arezoo.j88@gmail.com*

Abstract

Sport is a social phenomenon involving a large number of very different stakeholders. However, sport's front-line is occupied by sports organizations whose mission is to develop participation in sport and directly associated activities. With some knowledge, a few tips and some practice, sport organizations can quickly increase their ability to practice effective risk management. No single risk management model fits every organization. Different governance and administrative structures, and varying activities mean every organization must develop a risk management strategy to suit their specific needs. Every organization is exposed to risks in a variety of areas including governance, infrastructure, financial and operational risks.
In this paper, direct and indirect effects of risks are investigated with DEMATEL technique. Results show that governance risks have the greatest effect on sport organizations.

Keywords: DEMATEL, risk management, sport organization, effect.

INFORMATION SYSTEMS AND THE COMPETITIVE ADVANTAGE THEY CAN PROVIDE TO AFL CLUBS

Kelly Foreman [a b c]

[a] *University of South Australia*
[b] *Reconfigurable Computing Lab*
[c] *Corresponding author: Kelly.Foreman@unisa.edu.au*

Abstract

The use of GPS is becoming increasingly popular with coaches of elite Australian sporting clubs. To assist coaches with decision making relating to player performance, GPS equipment gathers data, such as players' speed; fatigue; heart rate; work rate; impact loading; reaction; effort; and the distance players run. However, the volume of information is problematic for coaches who are trying to interpret the data they are receiving and staff members who are trying to store the large volumes of data in easily accessible and meaningful formats. This is particularly evident given the tight deadlines between training sessions and matches during the season.

In this paper a system architecture is proposed for the design and installation of an information system that elite sporting clubs could use to store, maintain, access and report on the data they have collected from GPS units, the proposed architecture includes a data warehouse and decision-support system that will ensure data is easily stored and maintained, easily reported on and assists coaching staff with decision making based on GPS data.

This paper discusses the design, application and benefits of such a system in assisting coaching staff in making critical decisions about players based on reliable information that can be accessed within seconds of needing it through customized reporting.

Keywords: AFL, GPS, Information Systems, Decision-Support System, Competitive Advantage

---

1. INTRODUCTION

This paper will look at the current ways in which an AFL club is storing, organising and accessing different types of data provided to and/or generated within the football department and the issues with the current processes in place. A system architecture is then proposed that will streamline the current business and data storage processes creating customised reports and views for different staff requiring different data which will allow football department staff to spend more time with the players rather than hours of time copying, transferring and manipulating data to suit their individual needs.

2. THE CURRENT SITUATION

At the present time AFL clubs are battling with overwhelming amounts of information from a number of sources including GPS devices, statistics from third party companies, video, manual reporting from coaching staff, welfare statistics and information and the clubs own information gathering such as images and knowledge within the organisation.

At the present time the club on which this case study is based on is storing GPS data from each game and training session in individual spreadsheets. Exports of the PDF data are also kept, these are then stored in individual folders labelled in varying formats in which the data is manually reported on with a staff member opening individual files and writing the weekly GPS report based on those files, the weekly report also differing with content and formatting on a regular basis. The same format applies with GPS data from training and skills sessions as well as information about weights training. The files that come from the GPS units are not stored in any specific way and are scattered through multiple directories making it difficult to find the files needed to recreate the game and watch the player movements and data.

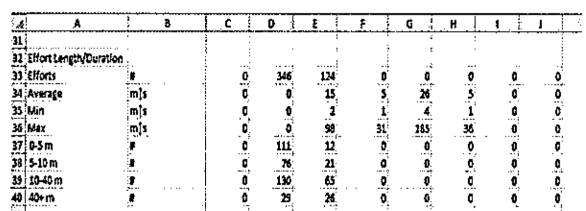
Player game day statistics are organised similarly in a directory based on the game, round and location, whilst player welfare information is stored in word documents.

The only shining light out of the way the club stores data is the use of injury management software provided by a third party company to enter injury, rehabilitation and doctors information into which is organised and stored in a database.

The way that data is being stored at the present time makes it difficult to compare individual player and team on field performances as well as providing an overwhelming amount of information to the end user. The exportation of GPS data to a spreadsheet is not only messy in the current way it is being done but is a time consuming process for the individual who has to perform the task for each GPS unit worn during the training session or game. Not only does it make a player comparison hard but it does not allow accurate and well informed decisions or comparisons to be made on player data which could affect the game plan, the clubs reliance on GPS devices or how a player is being treated for an injury or through the rehabilitation process. It also does not take into account the corruption of data that can easily occur when a spreadsheet reaches a certain file size or amount of cells filled with data.

Manual reporting based on this information is also time consuming as well as the format and information in the reports changing rather than being consistent every time it is produced.

The current processes in place by the club see everyone regardless of their position or need for specific data related to their area provided with the same data view. For example the fitness staff would see the same view of the data as the coaching staff even though both of them would be interested in different components of the data rather than having to scroll through pages of a spreadsheet trying to find the specific set of information or components they are interested in. An example of this can be seen in Figure 1.



	A	B	C	D	E	F	G	H	I	J
31										
32	Effort Length/Duration									
33	Efforts	#	0	346	124	0	0	0	0	0
34	Average	m/s	0	0	15	5	26	5	0	0
35	Min	m/s	0	0	2	1	4	1	0	0
36	Max	m/s	0	0	98	31	185	36	0	0
37	0-5 m	#	0	111	12	0	0	0	0	0
38	5-10 m	#	0	76	21	0	0	0	0	0
39	10-40 m	#	0	130	65	0	0	0	0	0
40	40+ m	#	0	29	26	0	0	0	0	0
41										
42	Start Accel.									
43	0-1 m/s²	#	0	80	23	0	0	0	0	0
44	1-2 m/s²	#	0	93	36	0	0	0	0	0
45	2-4 m/s²	#	0	118	39	0	0	0	0	0
46	4+ m/s²	#	0	49	26	0	0	0	0	0
47										
48	Recovery Times									
49	0-0.5 mins	#	318	318	58	318	318	318	318	318
50	0.5-1 mins	#	16	16	31	16	16	16	16	16
51	1-2 mins	#	3	3	21	3	3	3	3	3
52	2-5 mins	#	6	6	6	6	6	6	6	6
53	5+ mins	#	6	6	7	6	6	6	6	6
54										
55	Heart Rate									

Figure 1: An example of the GPS data exported directly from the manufacturer's software

3. PROPOSING A SYSTEM ARCHITECTURE

A better user interface and storage system would allow coaching and fitness staff to focus on their primary role of managing the players rather than spending time deleting copious amounts of cells from excel spreadsheets that are not needed because the coaching staff at that club don't look at those specific data attributes.

Business processes in combination with information systems are required in order to interpret, manage and action the data produced by GPS units, pre, post and during the game. Information systems can be used to automate existing business processes that are manually acted out by members of the organisation making them more efficient. This allows the organisation to make better decisions and improve the execution of their already existing business processes (Conway 2011). The Economist (2009, p.15) argues that "because technology under pins nearly every business process today, it can help those in the workplace improve their use of critical data". Therefore it is essential that AFL teams have structured business processes in place that allow information systems to leverage and automate them creating a more efficient use of time and resources for staff members.

In order to store the large amounts of GPS data generated each time a player wears the device it is recommended the data files be stored in a database or data warehouse in which data-driven decision support systems can be used to access the information and assist in decision making. This data warehouse will also have direct access to the injury management database as well as incorporate the storage, organisation and retrieval of video, game day statistics and other information critical to the football department. "These systems (data-driven support systems) analyse large pools of data found in major corporate systems. They support decision making by enabling users to extract useful information that was previously buried in large quantities of data" (Laudon and Laudon 2005, p.466). Data mining can also be used to analyse the data to assist coaching staff in making important decisions based on reliable information. This will not only ensure information from GPS units is stored in an easy to read and extract format but also that the data is useful to members of the organisation. It will also look for trends within the data stored in the databases.

The following system architecture is proposed which incorporates all of the aforementioned components. Figure 2 demonstrates a high level overview of the overall data warehouse structure and it's components.

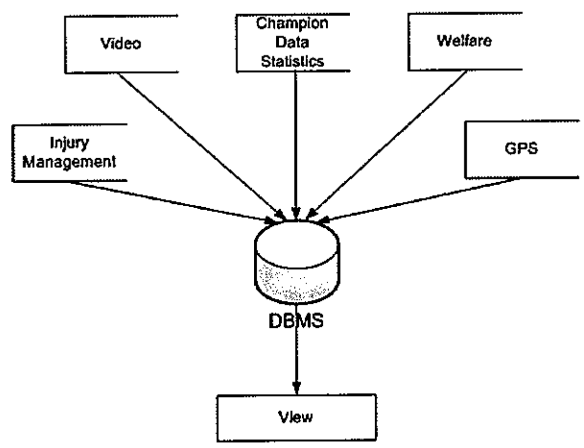


Figure 2: A high level view of the proposed system architecture

As seen in Figure 2 there is a central data warehouse, which is comprised of five data stores – injury management, video, champion data statistics, welfare and GPS. This leads to a customised view for the end user. For example if the midfield coach might only be interested in seeing champion data statistics from the weekends game view and GPS statistics but only for the midfielders, a simply query and report will present this information to them rather than them having to look through information for everyone in the team which are not involved in the line. This is in contrast to someone like the head fitness coach who may only be interested in the GPS statistics for every player on the ground in order to view their performance and modify training based on this. Therefore the unique view presented for each staff member minimises wasted time sifting through data and information they are not interested in.

Figure 3 demonstrates a more detailed technical view of the system architecture and how it works. The following steps take place when the end user runs a query regarding the information they require:

1. The client job process runs telling the database which view or report needs to be generated or accessed
2. The remote agent acknowledges the job and adds it to the queue
3. Remote agent executes the job command
4. The event processor starts the job

5. The event server looks for the next event to process
6. Job is complete
7. Display the results

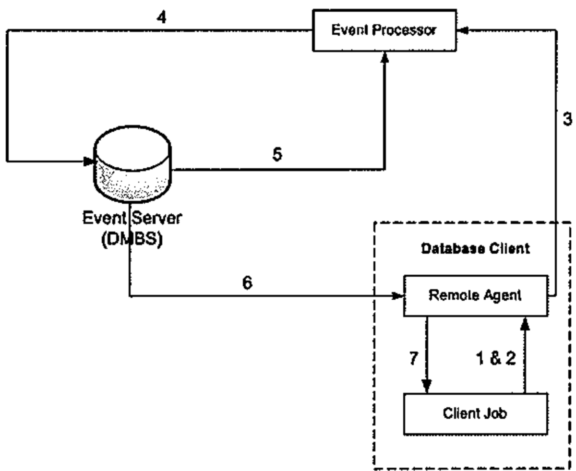


Figure 3: A low level view of the proposed system architecture

4. CONCLUSIONS

The incorrect storage of important data can see critical decisions regarding a players fitness or the teams overall strategy as and tactics be incorrectly made based on incorrect or inconclusive data that has been present in manual reporting. If a system architecture is implemented within the football department it will see streamlined business processes created and implemented which will automatically store, maintain and backup data as well as create customised views and reports for the individual needs of the staff in the football department which will allow them to spend more quality time with players rather than hours manually generating reports and entering data.

Acknowledgements

I wish to acknowledge the AFL clubs who participated in this study.

References

Conway, M. 2011. "Afl Hits New Mark in 2011." Retrieved 11 May, 2012, from http://www.afl.com.au/tabid/208/default.aspx?newsid=122668

Economist Intelligence Unit. 2009. "Organisational Agility: How Businesses Can Survive and Thrive in Turbulant Times," Economist Intelligence Unit, London, pp. 1-28.

Laudon, K., and Laudon, J. 2005. *Management Information Systems: Managing the Digital Firm*, (9th ed.). New Jersey: Prentice Hall.

APPLICATION OF GAME THEORY IN SPORT

Saeed Esyed Agha Banihashemi[a,b]

[a] *School of international relations of ministry of foreign affairs of Iran,*
[b] *Corresponding author: Ihusaied2001@yahoo.com*

Abstract

Game theory is a branch of mathematics which helps researchers in different subjects to choose best strategies to win. It is not only important in international relations, economics, the army, and social science, but in sport also, as in other subjects we have two players which each of them wishes to win. Also, other people wishing to bet on them in some sports like tennis, is subject to millions of dollars. Game theory can help them to choose which strategy is good for players, and it can be shown that game theory can predict the winner with 90% ability. In this article we explain theory and with some example show how is application of game theory in sport.

FAST TRACKING OF PLAYERS USING BACKGROUND SUBTRACTION AND COLLISION AVOIDANCE ALGORITHM

Hanyi Li [a,b], René C. Zorn [a], Urlich Rückert [a]

[a] *Cognitronics and Sensors System Group,*
Center of Excellence Cognitive Interaction Technology, Bielefeld University
[b] *Corresponding author: hli@cit-ec.uni-bielefeld.de*

Abstract

The Sports Performance Analyzer (SPA) is a system for analysis and evaluation of players' performance data. The SPA monitors players by using two cameras with fisheye lenses which are installed under the sports hall ceiling (Wilhelm et al., 2008). There is a high demand in sport and coaching science to get players' tracking data for real time scenarios quickly. Template matching as well as particle filter based trackers have been implemented in the SPA system (Monier et al., 2009, Monier, 2011). The frame rate of the trackers is not high enough to support the tracking in real time since the computational complexity is high. To solve this problem, we designed a new tracker in the SPA system, which uses a mixture model that incorporates information from both the dynamic model of each player and the detection achieved by background subtraction. To improve accuracy of the player detection pre- and post-processing of the image and the model is used. Our tracker validates foreground pixels by a moving player object model using both foreground and background statistics (Cheung & Kamath, 2005). A two-dimensional physical model of human motion is used to avoid tracking errors, which occur by non-elastic collisions when players come close to each other (Perše et al., 2005). The comparison of our tracker with the existing ones in the SPA system is described in the results. Some speed-up techniques of our implementation are shown as well in the paper.

Keywords: Video tracking, dynamic models, background subtraction

Acknowledgements

This work has been supported within the Excellence Initiative of the German Federal Ministry of Education and Research and the German Research Foundation (DFG) by the Cluster of Excellence "Cognitive Interaction Technology" (EXC277).

References

Wilhelm, P., Monier, E., Xu, F., & Witkowski, U. (2008). Analysis of Indoor Team Sports Using Video Tracking and Wireless Sensor Network. *Current trends in Performance Analysis: World Congress of Performance Analysis of Sport VIII*, 345-348.

Monier, E., Wilhelm, P., & Ruckert, U. (2009). Template Matching Based Tracking of Players in Indoor Team Sports. *Third ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC 2009)*, Como, Italy, 30 – 2 Sep 2009.

Monier, E., (2011): Vision Based Tracking in Team Sports. Heinz Nixdorf Institute, University of Paderborn, PhD Thesis.

Perše, M., Perš, J., Kristan, M., Vu kovi , G., & Kova- i , S. (2005). Physics-based modelling of human motion using kalman filter and collision avoidance algorithm. *International Symposium on Image and Signal Processing and Analysis, ISPA05*, Zagreb, Croatia, 328-333, Sep 2005.

Cheung, S., & Kamath, C. (2005). Robust Background Subtraction with Foreground Validation for Urban Traffic Video. *EURASIP Journal on Applied Signal Processing, Volume 14*, 1-11, 2005.