Proceedings of the 13th Australasian Conference on

# MATHEMATICS AND COMPUTERS IN SPORT

## Melbourne, 11-13 July, 2016

Edited by Ray Stefani and Adrian Schembri


Shane Warne


Cathy Freeman

# THE PROCEEDINGS OF THE 13TH AUSTRALASIAN CONFERENCE ON MATHEMATICS AND COMPUTERS IN SPORT

Edited by

Ray Stefani and Adrian Schembri

JULY 11th to 13th, 2016

MELBOURNE, VICTORIA, AUSTRALIA

Published by ANZIAM MathSport

ISBN: 978-0-646-95741-8

*The papers and abstracts in these proceedings have gone though a full peer review refereeing process.*

# ANZIAM Mathsport 2016

The biennial Mathsport conference is proud to return to Melbourne for 2016. The City Conference Center venue of Victoria University was purposely chosen, given that it is close to the Melbourne Cricket Ground and to Melbourne Park with its myriad of sporting venues. Arguably, Melbourne could host a Summer Olympics with just the existing venues, with much of the activity near the City Conference Center.

In addition to the scientific events, we have several social events planned that will take advantage of Melbourne’s great dining and sporting venues. On 10 July, the day before we start, there is a footy match between Carlton and Adelaide at the MCG. We will visit the National Sports Museum at the MCG on 11 July after day one. Our conference dinner on 12 July is at nearby Young and Jackson pub at Flinders and Swanston.

A major goal of the Mathsport conference is to provide a stage for leading thinkers in sports analysis in the Asia-Pacific to share some of their recent work and how it is making an impact in sport. This proceedings includes information about the 39 talks to be given at Mathsport 2016: three keynote talks, two guest talks, six talks on the AFL, five talks on analytic methodologies of general application, five on cricket, four on gambling, three on rugby, two on golf, two on tennis and one each on seven other areas.

On behalf of my Organizing Committee colleagues Stephanie Kovalchic, Denny Meyer, Sam Robertson and Adrian Schembri, we look forward to hearing the presentations and we hope the attendees and readers of these proceedings will come away with many new ideas.

Ray Stefani

# Table of Contents

## Olympics- Ancient Games



## Rugby



## Soccer



## Tennis

## Keynote Speakers

### Gary Mc Coy

Gary and Mathsport founder Neville de Mestre, are working to develop a uniform metric for surf safety. The goal is to measure the level of danger in surfing competition as well as for recreational surfing activities.

### Ken Quarrie

Ken is Senior Scientist for the All Blacks.  NZ Rugby employed him as their inaugural Injury Prevention Manager in 2000. The purpose of this role was to put in place an evidence assisted, nationwide injury prevention program for New Zealand rugby, with a special emphasis on the prevention of the most severe injuries. Subsequent to the introduction of the program, entitled RugbySmart, there has been a reduction of over 90% in the annual rate of scrum-related spinal injuries, and a reduction of over 50 over all permanently disabling injuries.

### Sam Robertson

Sam wears many hats. He is Senior Sport Scientist for the Western Bulldogs, Senior Research Fellow at Victoria University and he provides research and innovation for Golf Australia.

## Guest Speakers

### Liam Lenten

Liam is a Senior Lecturer in Applied Econometrics at the Department of Economics and Finance at La Trobe University, Melbourne. He has been there since 1997. Liam earned his honours Economics degree there in 1995; and then a Master of Commerce degree at the University of Melbourne. Liam then undertook his PhD thesis from 1999-2005, which highlighted his interest in exchange rate determination models and macroeconomic cycles. However, his research more recently has centred more on sports economics, specifically rules, regulation, incentives and athlete behaviour; with an emphasis on various forms of cheating (doping, match-fixing, etc.). Liam has held visiting positions at: University of Michigan (US); Massachusetts Institute of Technology (US); University of Otago (NZ); Lancaster University (UK); University of Exeter (UK) and Monash University (Aust). Liam has published 27 articles in peer-reviewed journals, including *European Journal of Operational Research, Sport Management Review*, *Journal of Sports Economics*, *Journal of Forecasting*, *Australian Journal of Management* and *Applied Economics*.

### Western Bulldogs

Mathsport will welcome coaching and playing representatives from the Bulldogs.

# A STATISTICAL APPROACH TO ENHANCING PLAYER SKILL IN AUSTRALIAN RULES FOOTBALL: APPLICATION TO TEAM SPORT

Sam Robertson [a,b]

[a] *Institute of Sport, Exercise & Active Living (ISEAL), Victoria University*
[b] *Western Bulldogs AFL Club*

*Corresponding author:* sam.robertson@vu.edu.au

## Abstract

Principles of learning state that the specificity, progression and variety of skill training undertaken by an individual are associated with the level of improvement expected (Henry, 1968; Pinder et al., 2011). Despite this, the manner by which these principles are quantified in the field from a skill acquisition perspective is not well-established (Farrow & Robertson, 2016). This presentation provides an overview of the existing skill acquisition program at the Western Bulldogs AFL club. Specifically, it details how through blending traditional performance analysis, statistics and machine learning, interventions targeting the skilled behaviour of players have been implemented and evaluated. One example outlines the use of rule induction techniques to understand the dynamic interaction between task and environmental constraints in influencing player skill execution during matches and training (Newell, 1986). A second reveals the application of statistical process control approaches to monitoring player performance and progression. A third shows illustrations of modelling expected rates of development of players based on characteristics relating to their selection in the AFL Draft.
Concepts worthy of future research are presented, including methods for integrating skill- and physical-related data, quantifying the contribution of the individual to team outcomes and optimising player rotation strategies.

*Keywords*: Analytics, AFL, Team Sport

### Acknowledgements

I wish to thank the Victoria University – Western Bulldogs PhD & Honours students and cadets that contributed and continue to contribute to making this work possible.

### References

Henry, F. M. (1968). Specificity vs. generality in learning motor skill. Classical studies on physical activity, 328-331.

Newell, K. M. (1986). Constraints on the development of coordination. *Motor development in children: Aspects of coordination and control, 34*, 341-360.

Pinder, R. A., Davids, K. W., Renshaw, I., & Araújo, D. (2011). Representative learning design and functionality of research and practice in sport. *Journal of Sport and Exercise Psychology, 33*, 146-155.

Farrow, D., & Robertson, S. (2016). Development of a skill acquisition periodisation framework for high-performance sport. In-Press.

# DEVELOPING A CLASSIFICATION OF PLAYER ROLES IN AUSTRALIAN RULES FOOTBALL

Adrian Barake, Heather Mitchell, Constantino Stavros & Mark Stewart

*RMIT University*
*School of Economics Finance & Marketing*
*Corresponding author: adrian.barake@rmit.edu.au*

**ABSTRACT**

Australian rules football is an ever-evolving collision sport, based on the principle of territorial invasion. The nature of the game, coupled with both a large playing field and a high number of participants requires athletes with a range of increasingly dynamic and multi-faceted skills. Players are required to contest possession, and both attack and defend. Historically, the role, or position, of players has been linked to the location on the ground that the player primarily occupies. Champion Data (CD), the official supplier of statistics to the Australian Football League (AFL), currently utilizes seven 'positions' for classifying players: Key Defender, General Defender, Midfielder, Midfielder-Forward, Key Forward, General Forward and Ruckman. CD assigns players to these positions on a season-by-season basis, rather than a match-by-match basis; with the classification based on subjective assessment. The result of this approach is that for those players who have played in multiple positions (and there are an increasing number of such players), performance assessment is problematic as the performance indicators for each of these seven positions differ markedly. This research utilizes match statistics captured across each zone of the ground to perform a multi-nomial logistic regression to allocate a position to every player in each match. This approach enabled the classification process to be performed on not only AFL games, but also second tier football games for which CD also collates statistics. An ability to classify players on a match by match basis will greatly enhance the value of the player statistics collected. This will be particularly important to AFL recruiters when applied to secondary competitions. In addition to mapping a player to existing positional classifications, a new segmentation of player positions based on a combination of player location and playing styles is being developed, extending the number of positional types.

*Keywords*: Australian football, sports statistics, player classification

## 1. INTRODUCTION

The game of Australian football is a territorial invasion sport, played on a relatively large field with thirty-six on field participants. It has no off-side rule or other such restriction on a player's movement during the course of a match. This produces a fast moving game played by athletes required to have multi-dimensional skills and ever increasing fitness capabilities. Each player is required to contest for the ball, defend against opposition scoring attempts and distribute the ball effectively when in possession. Whilst a player invariably has a starting position and an identifiable role during the course of play, this is dictated almost exclusively by the strategy and tactics of a team - not the rules that govern the game. There is currently no objective method for identifying the position of a player in each match, and this paper seeks to address this gap through the application of a multinomial logistic regression model.

### 1.1 Australian football position classifications

Historically, the position of players in Australian football has been linked to the location on the ground that each of the 18 players primarily occupies. For many years the game was played in a manner where the starting position of a player would dictate the primary area on the ground that a player would occupy. However, more recent advances, including increased player fitness and evolving coaching tactics has brought about changes to the location and position of players, such that they are now universally accepted as being markedly different from the traditional position classifications. Champion Data (CD), the official supplier of statistics to the Australian Football League (AFL), currently utilizes seven positions for classifying players: Key Defender, General Defender, Midfielder, Midfielder-Forward, Key Forward, General Forward and Ruckman.

## 1.2 Match position classifications

CD currently assigns players to positions on a season-by-season basis, rather than a match-by-match basis. This classification is based on a subjective assessment of a player's statistics over the course of a season. The key performance indicators for a player are linked to the position of a player. Whilst all players on the ground are expected to contest for possession, defend and attack as required, there are distinctions in the statistical profile of players in different positions. As a result, to make a meaningful assessment of a player's contribution it is relevant to know the position of the player in that match (or group of matches). For players who are used in the same position in each match this is not an issue, but for the increasing number of players who are required to play in different positions, performance assessment is problematic.

## 1.3 Implications

The AFL operates under the equalisation measures of salary caps and player drafts, which places greater importance on the recruitment of new and/or junior players into the competition. The assessment of such players is also shaped by the position of a player, and as such this paper seeks to provide a method for segmenting players (across all competitions covered by CD) on a match-by-match basis into the seven CD positions.

## 1.4 Multinomial logistic regression

The allocation of players to positions is a statistical classification problem, with a nominal dependent variable for which there are more than two categories (i.e. the position outcomes). Multinomial logistic regression is a classification method that is used to predict possible outcomes of such a categorically distributed dependent variable, with the independent or explanatory variables in this scenario being the match statistics collated for each player. One of the benefits in using a multinomial logistic regression is that the model will provide probability estimates for each of the possible outcomes, rather than just a binary prediction. In this case where each of the possible outcomes represents a different position, a probability estimate provides both an indication of certainty and a means for determining secondary hybrid positions.

# 2. METHODS

All of the match statistics collated by CD are segmented into the zone on the ground in which they occurred. The three different zones CD utilises for all competitions are: defensive 50 (metres from goal being defended), forward 50 (metres from goals being attacked) and midfield (the remainder of the ground). This segmentation of match statistics by zone provides a starting point for identifying a player's position during a match. The ultimate position estimate for a player in this paper is a two-step approach. Firstly, the multinomial logit model segments players into locations, and then a binary logit model classifies players by type (*Key* or *General).* This two-step approach can be done either as a binary or multiplicative process.

## 2.1 Model alignment with CD positional classifications

CD engages staff at all AFL games to capture a variety of performance metrics, including the time spent by players in various locations throughout the game. This results in a *Time-In-Position* report that segments a player's time into four categories: Defence, Midfield, Forward and Ruck. These categories are similar to the three ground zones, with the additional category of Ruck. The *Ruckman* is a roaming role that is not defined by zone, and is the most easily distinguishable position. The *Ruckman* contests for the ball following each stoppage in the game by attempting to hit/palm the ball to a team-mate in the manner of a 'tip-off' that commences a basketball game. These actions are recorded as *Hit-outs* and are the key variable in categorising a player as a *Ruckman*. The four categories from the *Time-In-Position* report have been used to represent the four possible outcomes (dependent variable) in the multinomial regression model. These four primary positions can then be used as the basis for further segmentation into the seven CD positions (the *Mid-Forward* position can be derived from the regression model as a supplementary component – similar to CD's current classification of players who have 40%, or more, of game time in each zone).

The distinction between *Key* or *General* defenders/forwards is currently a subjective allocation. The considerations for this classification include a player's height and the balance of aerial and ground involvements i.e. a *Key* position player will generally be taller and have more marks/spoils compared to a *General* player. For this paper the classification of players as a *Key* or *General* type player has been derived from a separate logit regression model, and then held as a constant across a season as there would often not be an explicit intent for a change in player type from match to match.

### 2.2 Variables and data included in the test models

The 2015 AFL season was used as the test data for assigning match positions, but as the model is required to predict positions of players in second tier competitions, only the types of match statistics captured in these secondary competitions could be considered. In addition to the statistics captured by CD, some basic derived variables were created that would be expected to have strong correlations with specific positions. The three such variables that were eventually included in the ultimate models were:

- *Score impact to possessions*: the ratio of shots at goal and score assists combined compared to possessions.
- *Ground level possessions*: simply possessions less marks.
- *Mark rate*: the number of marks relative to the number of possessions.

The statistics included in the models (and the naming conventions) are detailed in Table 1 below.

| Statistic | Abbrev |
|---|---|
| Tackle | tack |
| Spoil | spoil |
| Smother | smoth |
| Bounce | bounce |
| Mark: from opposition kick | mk_fok |
| Mark: contested | mk_con |
| Mark | mk |
| Clearance | clr |
| Hit-out | hit |
| Possession | poss |
| Shot at goal | sag |
| Score impact relative to possessions | score_poss |
| Marks relative to possessions | mk_rate |
| Ground level possessions | glvl |
| Player's height | height |

| Description | Notation |
|---|---|
| Tackle in the defensive 50 | d_tack |
| Tackle in the midfield | m_tack |
| Tackle in the forward 50 | f_tack |
| Total tackles | t_tack |
| Season average tackles | s_tack |

As the model developed is required to be used at second tier level - which includes junior players - the height variable was standardised to improve the efficacy in extrapolating the model to other competitions.

Table 1: Match statistics used for player position assignment and notations

### 2.3 Application of multinomial logistic regression

The multinomial logit model is used to provide a prediction, or probability estimate, that a player was in each of the four positions. The base outcome (in this model a *Defender*) is derived from equation (1), and the outcomes for all other positions is derived from equation (2).

$$\mathrm{P}(y_i = 1 \mid x_i) = \frac{1}{\left(1 + \sum_{k=2}^{J} e^{x_i'\beta_k}\right)} \qquad (1)$$

$$\mathrm{P}(y_i = j \mid x_i) = \frac{e^{x_i'\beta_j}}{\left(1 + \sum_{k=2}^{J} e^{x_i'\beta_k}\right)} \qquad (2)$$

Where $x_i'$ is the set of match statistics for player $i$,
$\beta_k$ is a vector of regression coefficients (Table 3),
$j$ is each position in a set of $J$ outcomes,
$k$ is the number of levels of the dependent variable where; $j \neq k$ = 1, 2, 3, 4 and $J$ = 4,
$\beta_j$ is the regression coefficients for position $j$.

### 2.4 Logistic regression for player type

The logistic model for classifying defenders/forwards as *Key* or *General* is similar to the multinomial logit model, but is a binary model and as such has one set of regression coefficients and just the one prediction.

$$\mathrm{P}(y_i = 1 \mid x_i) = \frac{e^{x_i'\beta_1}}{\left(1 + e^{x_i'\beta_1}\right)} \qquad (3)$$

For the purposes of this research, a *Ruckman* is also classified as being a *Key* position player, as there were a number of players who alternated between *Ruck* and *Key Forward/Defender*, but none who alternated between *Ruck* and *General Forward/Defender*.

The functions of a *Key Defender* are different from those of a *Key Forward*¸ and as a result two different models are used for classifying a player as *Key* or *General*, with the main distinction being the logit model for a *Key Defender* references spoils.

## 3. RESULTS

The correlations between 2015 AFL match statistics in each zone and the player positions as assessed by CD are shown in Table 2. Not surprisingly the strongest correlations for a defender were with statistics captured in the defensive 50, and vice versa for a forward. It was also evident that there was a positive correlation between a defender and some statistics from the midfield zone – *Spoils* and *Marks from opposition kicks*.

In determining which match statistics to include in the model as independent variables, a combination of those statistics that had the strongest and significant correlations, as well as independent actions were trialled. A number of the statistics that are captured are strongly correlated e.g. shots at goal and goals scored. Selecting statistics conceptually and statistically independent helped to manage the potential issue of multi-collinearity.

| **Statistic** | **def** | | **mid** | | **fwd** | | **ruck** | |
|---|---|---|---|---|---|---|---|---|
| z_height | 0.096 | *** | (0.286) | *** | (0.033) | *** | 0.457 | *** |
| d_tack | 0.262 | *** | 0.033 | *** | (0.273) | *** | (0.049) | *** |
| d_spoil | 0.519 | *** | (0.254) | *** | (0.274) | *** | 0.005 | |
| d_bounce | 0.186 | *** | (0.061) | *** | (0.103) | *** | (0.049) | *** |
| d_poss | 0.567 | *** | (0.015) | | (0.493) | *** | (0.133) | *** |
| m_mk_fok | 0.358 | *** | (0.141) | *** | (0.233) | *** | 0.023 | ** |
| m_spoil | 0.390 | *** | (0.263) | *** | (0.190) | *** | 0.118 | *** |
| m_tack | (0.301) | *** | 0.383 | *** | (0.104) | *** | 0.057 | *** |
| m_glvl | (0.233) | *** | 0.572 | *** | (0.299) | *** | (0.068) | *** |
| f_tack | (0.346) | *** | 0.074 | *** | 0.281 | *** | (0.011) | |
| f_spoil | (0.139) | *** | (0.082) | *** | 0.198 | *** | 0.051 | *** |
| f_smoth | (0.132) | *** | 0.020 | * | 0.119 | *** | (0.013) | |
| f_poss | (0.525) | *** | 0.050 | *** | 0.504 | *** | (0.052) | *** |
| t_clr | (0.386) | *** | 0.554 | *** | (0.213) | *** | 0.111 | *** |
| t_hit | (0.185) | *** | (0.142) | *** | (0.092) | *** | 0.873 | *** |
| t_sag | (0.413) | *** | (0.054) | *** | 0.498 | *** | (0.058) | *** |
| score_poss | (0.432) | *** | (0.143) | *** | 0.584 | *** | (0.013) | |

*** p<0.01, ** p<0.05, * p<0.10

Table 2: Correlations between zone match statistics and 2015 AFL player positions

The coefficients for the multinomial logit model are shown in Table 3, with all of the variables being significant at the 1% level for at least one position, except for forward spoils (*f_spoil*) and total shots on goal (*t_sag*) which are both significant at the 5% level for the *Forward* position (Table 3).

| **Postion →** | **Mid** | | | **Fwd** | | | **Ruck** | | |
|---|---|---|---|---|---|---|---|---|---|
| **Ind. Var. ↓** | Coef. | Std. Err. | | Coef. | Std. Err. | | Coef. | Std. Err. | |
| z_height | (0.070) | 0.075 | | (0.247) | 0.079 | *** | 1.639 | 0.307 | *** |
| d_tack | (0.397) | 0.059 | *** | (0.700) | 0.074 | *** | (0.788) | 0.261 | *** |
| d_spoil | (0.981) | 0.086 | *** | (1.078) | 0.106 | *** | (1.047) | 0.196 | *** |
| d_bounce | (0.393) | 0.137 | *** | (0.062) | 0.173 | | (3.513) | 1.798 | * |
| d_poss | (0.239) | 0.021 | *** | (0.567) | 0.028 | *** | (0.558) | 0.105 | *** |
| m_mk_fok | (0.630) | 0.083 | *** | (0.897) | 0.106 | *** | (0.578) | 0.282 | ** |
| m_spoil | (0.718) | 0.065 | *** | (0.472) | 0.069 | *** | (0.698) | 0.132 | *** |
| m_tack | 0.258 | 0.036 | *** | 0.133 | 0.039 | *** | 0.253 | 0.090 | *** |
| m_glvl | 0.069 | 0.013 | *** | (0.079) | 0.015 | *** | (0.046) | 0.053 | |
| t_clr | 0.678 | 0.047 | *** | 0.298 | 0.052 | *** | 0.782 | 0.113 | *** |
| t_hit | 0.390 | 0.062 | *** | 0.522 | 0.063 | *** | 0.766 | 0.067 | *** |
| f_tack | 0.740 | 0.104 | *** | 0.967 | 0.107 | *** | 0.662 | 0.204 | *** |
| f_spoil | 0.075 | 0.226 | | 0.475 | 0.229 | ** | 0.447 | 0.351 | |
| f_smoth | 1.043 | 0.301 | *** | 1.515 | 0.318 | *** | 1.146 | 0.657 | * |
| f_poss | 0.698 | 0.060 | *** | 0.991 | 0.063 | *** | 0.769 | 0.132 | *** |
| t_sag | 0.055 | 0.096 | | 0.251 | 0.100 | ** | (0.359) | 0.220 | |
| score_poss | 1.996 | 0.895 | ** | 3.723 | 0.884 | *** | 3.598 | 1.640 | ** |
| _cons | (0.781) | 0.144 | *** | 0.746 | 0.146 | *** | (5.564) | 0.700 | *** |

*** p<0.01, ** p<0.05, * p<0.10

Table 3: Match position regressions

The accuracy of the multinomial model in terms of using match by match data to correctly allocate players to their CD time in position location is shown in Table 4. A total of 9,060 player matches were assessed and 83.91% of the positions with the highest probability align with CD's time in position assessments. The model

| 2015 Prediction | Actual position (per CD Time-In-Position) Def | Mid | Fwd | Ruck | Total |
|---|---|---|---|---|---|
| True positives | 91.52% | 76.19% | 82.43% | 90.59% | 83.91% |
| False positives | 10.12% | 21.05% | 18.65% | 10.35% | |

was able to accurately predict more than 90% of games played by *Defenders* or *Ruckman*, but was not as accurate in predicting *Midfielders* (76%) and *Forwards* (82%). This is partly due to the interchangeability between some midfielders and forwards that has been recognised by CD with its hybrid *Mid-Forward* position.

Table 4: Multinomial model confusion matrix for 2015 AFL data

The generalisability of the model has been tested on the 2014 AFL season. Using the same logic of the maximum position estimate being the final position predicted, the model was able to accurately predict 82.77% of positions, with similar results on a positional basis, i.e. *Defenders* and *Ruckman* were more accurately predicted. These results indicate that the predictive power of the model is transferrable to other AFL seasons.

| 2014 Prediction | Actual position (per CD Time-In-Position) Def | Mid | Fwd | Ruck | Total |
|---|---|---|---|---|---|
| True positives | 90.48% | 76.30% | 80.09% | 89.69% | 82.77% |
| False positives | 11.07% | 21.96% | 20.22% | 10.49% | |

Table 5: Multinomial model confusion matrix for 2014 AFL data

### 3.1 Player type

A similar process for establishing whether a player was a *Key* or *General* type player was followed, with the first step being to assess the correlations between match (season) statistics and CD season player classifications for 2015. As this research was holding a player's type constant, the match statistics for players were averaged across a season (and each variable is prefixed with '*s*'). The correlations for the variables trialled in the models are included in Table 6, and as expected the correlation between *s_spoil* and *Key Defender* (0.71) is stronger than that for *Key Forward* (0.55). The only variables that were included in the models were those that were conceptually relevant, avoided issues with multicollinearity (e.g. including two variables linked to spoils*)*, and minimised model complexity. The final binary logit model for classifying a *Key Forward* included a player's height, season averages for possessions and contested marks, along with total marks relative to possessions. The model for determining a *Key Defender* used all of the same variables with season average spoils also included.

| Statistic | key_def | | key_fwd | |
|---|---|---|---|---|
| z_height | 0.717 | *** | 0.806 | *** |
| s_mark | 0.115 | *** | 0.431 | *** |
| s_poss | (0.342) | *** | (0.201) | *** |
| s_mk_con | 0.365 | *** | 0.684 | *** |
| s_spoil | 0.709 | *** | 0.554 | *** |
| s_mark_rate | 0.437 | *** | 0.624 | *** |
| s_spoil_rate | 0.674 | *** | 0.476 | *** |

*** p<0.01, ** p<0.05, * p<0.10

Table 6: Correlations between season statistics and player type

The accuracy of the binary logit models for predicting whether a player was a *Key* or *General* type player are shown in Table 7. For both models a player was deemed to be a *Key* player type if the probability estimate was greater than 0.5. The model for *Forwards* was able to accurately predict 97.32% of the 224 forwards as either *Key* or *General* type players. The model for *Defenders* was able to accurately predict 91.89% of the 222 players classified as a defender as either *Key* or *General* type players.

| 2015 Prediction | Actual Forwards Key (K) | Gen (~K) | Total |
|---|---|---|---|
| Key (+) | 73 | 5 | 78 |
| General (-) | 1 | 145 | 146 |
| Total | 74 | 150 | 224 |
| Sensitivity | | Pr ( + \| K ) | 98.65% |
| Specificity | | Pr ( - \| ~K) | 96.67% |
| Positive Predictive Value | | Pr ( K \| + ) | 93.59% |
| Negative Predictive Value | | Pr ( ~K \| - ) | 99.32% |
| Correctly classified | | | 97.32% |

| 2015 Prediction | Actual Defenders Key (K) | Gen (~K) | Total |
|---|---|---|---|
| Key (+) | 72 | 9 | 81 |
| General (-) | 9 | 132 | 141 |
| Total | 81 | 141 | 222 |
| Sensitivity | | Pr ( + \| K ) | 88.89% |
| Specificity | | Pr ( - \| ~K) | 93.62% |
| Positive Predictive Value | | Pr ( K \| + ) | 88.89% |
| Negative Predictive Value | | Pr ( ~K \| - ) | 93.62% |
| Correctly classified | | | 91.89% |

Table 7: Logit model confusion matrix for 2015 AFL data for player type

When the models were applied to the 2014 AFL season, the model for *Forwards* was able to accurately predict 94% of the 238 forwards as either *Key* or *General* type players. The model for *Defenders* was able to accurately predict 91% of the 215 players classified as a defender as either *Key* or *General* type players.

## 4. DISCUSSION

As expected, the greater the number of events a player had in a game that were included in the model, the greater the accuracy in the prediction of that player's position. The final step in the development of the model is to incorporate prior information, as the prior position of a player is a good indicator of current position. A number of different methods have been provisionally trialled to incorporate such prior information (where available) into the prediction of current position. These include incorporating lags into the multinomial logit model, which improved the model's predictive ability but increased complexity, a Bayesian approach which at first run was too slow to identify positional changes, and using a Poisson distribution based on the number of match events to determine the weighting given to the current and prior match information.

### 4.1 New positional classifications

The probability estimates provided by the multinomial logit model offer scope for the creation of new position cohorts. This includes additional secondary positions such as *Ruckman-Forward* and *Defender-Mid.* The logit models for player type provide opportunity for creating a third player type *Hybrid*, that would recognise those players who are at the margins of being classified *Key* or *General*, and practically do play in either position.

### 4.2 Research limitations

Consideration was given to weighting match statistics to account for team strength – a player in a high scoring team will have more opportunities to have forward 50 involvements, and a player in a struggling team will have more opportunities to have involvements in the defensive 50. However, the marginal improvement in the predictive ability of the models tested was thought to not outweigh the complexity and subjectivity that this method introduced. Ideally match statistics would be standardised on a per minute basis, but this information is not available for secondary competitions, so could not be included in the model.

## 5. CONCLUSION

This paper provides a method for classifying Australian football players into positions on a match-by-match basis. The multinomial logit model developed was able to accurately predict the position of a player in more than 80% of 9,060 player games, with similar results achieved when extrapolating the model to other AFL seasons. The ability to segment every game for a player by position has the potential to improve player assessment at both AFL level, and critically second tier competitions where players are recruited into the AFL. The requirement of players to perform in multiple positions requires a performance analysis framework that can segment player games into multiple positions. The model detailed in this paper provides the methodology for such a classification in Australian football.

**Acknowledgements**

We wish to thank Champion Data for its assistance in providing match statistics.

**References**

Champion Data (2016). *AFL Prospectus: The essential number cruncher for season 2016.* Southbank Australia.

Coates, D & Oguntimein, B 2010, 'The Length and Success of NBA Careers: Does College Production Predict Professional Outcomes?', International Journal of Sport Finance, vol. 5, no. 1, pp. 4-26.

Hughes, M, Caudrelier, T, James, N, Donnelly, I, Kirkbride, A & Duschesne, C 2012, 'Moneyball and soccer - an analysis of the key performance indicators of elite male soccer players by position', Journal of Human Sport & Exercise, vol. 7, pp. 402-12.

Pyne, DB, Gardner, AS, Sheehan, K & Hopkins, WG 2006, 'Positional differences in fitness and anthropometric characteristics in Australian football', Journal of Science and Medicine in Sport, vol. 9, no. 1–2, pp. 143-50.

Tiedemann, T, Francksen, T & Latacz-Lohmann, U 2011, 'Assessing the performance of German Bundesliga football players: a non-parametric metafrontier approach', Central European Journal of Operations Research, vol. 19, no. 4, pp. 571-587.

Wooldridge, J. M. (2010) Econometric Analysis of Cross Section and Panel Data, 2nd Edition, MIT Press.

# MEASURING THE SIMILARITY BETWEEN PLAYERS IN AUSTRALIAN FOOTBALL

Jackson, Karl [a]
[a] *Champion Data – Melbourne, Australia:* karl.jackson@championdata.com.au

**Abstract**

This paper introduces a measure of similarity between players in the Australian Football League (AFL). For each player, their signature is defined using a 175-dimension vector that represents a breakdown of a player's involvements during games. Each dimension is a unique combination of event type (43 classifications), event zone (three classifications) and game state (three classifications). The magnitude of each dimension is defined as the percentage of a player's total involvements that were observed for that particular combination of event, zone and state. The similarity between two players is then taken as a linear transformation of the vector angle between players. The two most-similar players during the 2015 AFL season were found to be key forwards Josh J. Kennedy (West Coast Eagles) and Jeremy Cameron (GWS Giants). This similarity measure will also be used to graphically represent a squad's playing list, to identify unique players, and to match players from other leagues (such as new draftees, and women's football) to their AFL equivalents.

**Keywords**: Similarity, Australian Football, Classification

## 1. INTRODUCTION

Australian Football is competed at the highest level in the Australian Football League (AFL). As of the 2016 season, this league is contested by 18 clubs spread over five of Australia's eight states and territories. Each of Western Australia, South Australia, New South Wales and Queensland are represented with two teams, with 10 being based in Victoria, nine of those in Melbourne. Each club has a playing roster of roughly 45 players, with 22 to be selected to compete in each game – 18 starting on the field and four as interchange players. Clubs can then use up to 90 player rotations within games to interchange players. There are no restrictions on positions once on the field and players are often expected to perform in multiple roles not only across the course of a season, but also within games.

AFL clubs can add players to their roster between seasons by either trading players and/or draft picks with other AFL teams, signing free agents from other AFL teams or selecting new players via an end-of-season draft. Table 1 gives a summary of the new players selected via the draft for the 2016 AFL season. The vast majority came from state league competitions (80%) while the remainder were a mixture between lower-grade competitions and players recruited from other sports such as basketball and Gaelic football.

| Competition | State | 2015 Players | Players Drafted | Draft % | % of New Players |
|---|---|---|---|---|---|
| TAC Cup | Victoria | 773 | 45 | 5.8% | 39% |
| SANFL | South Australia | 1045 | 16 | 1.5% | 14% |
| WAFL | Western Australia | 1149 | 13 | 1.1% | 11% |
| NEAFL | QLD, NSW, ACT, NT | 492 | 9 | 1.8% | 7.8% |
| VFL | Victoria | 755 | 8 | 1.1% | 7.0% |
| Other | | N/A | 24 | N/A | 21% |
| Total | (Excluding "Other") | 4077 | 91 | 2.2% | 79% |

Table 1: Source leagues for new 2016 AFL players.

Club recruiters spend their time throughout the season narrowing a potential list of thousands of players to a draft board of the order of 100 players from which they will select 6-10 to recruit. This process involves a large amount of time-consuming manual work, including identifying strengths and weaknesses of players. Significant time could be saved with an accurate indicator of a player's playing style, which we will attempt to address in this paper through the derivation of 'similar players'.

There have been several previous attempts to cluster players into predefined positions, such as Sargent & Bedford (2010) who used 13 event types to place players into four categories – defenders, midfielders, forwards and ruckmen – and Pyne et al (2006) who used fitness testing and physical characteristics to predict a player's future position at AFL level. Though Sargent & Bedford (2010) did introduce a concept of dissimilarity between players, this concept of having a continuous representation of player roles has not been rigorously developed in team sports.

Champion Data maintains position labels for players, but these are restricted to seven categories:

- Key Defender
- General Defender
- Ruckman
- Midfielder
- Mid-Forward
- General Forward
- Key Forward

These positions can essentially be reduced to identifying which end of the ground a player spends the majority of their time (forward, midfield or defence) and their height (key or general, rucks). Introducing a more granular representation of a player's style with more position types should enable better, and faster, identification of player strengths and weaknesses, since recruiters would be prompted to a player's style before seeing that player play a game. Identifying players similar to a particular player would also enable smarter roster-management decisions by AFL clubs, where they can avoid having players with two closely-related skill sets on the same team, or find targets to potentially replace an outgoing player. Members of the general public, through coverage in the media, could also better identify with young players if they are identified as similar to a more senior player with a similar playing style.

## 2. METHODS

A player's involvement in games was measured as a combination of event type (how the player was involved), the current state of the game, and the location of the event.

Champion Data records more than 100 event types for AFL games. For this paper 43 of these events will be used as a representation of a player's involvement in games. Events that were excluded include those that happen irregularly, such as dropped marks, and events that are only recorded at AFL level and not at lower levels, like ruck contests. Many event types also take into account the state of the game – whether a player's teammate was in control of the ball, the opposition was in control or the ball was in dispute. These classifications were removed to further reduce the number of events recorded.

Though the game state was removed from the event type, it was reintroduced with slightly altered logic as a second variable. Game state was defined as:

1. Interception — Winning the ball off the opposition;
2. Stoppage — Involvements at stoppages before either team has cleared the area;
3. General Play — All other involvements.

In this manner, a player taking a mark off an opposition kick is treated differently to a mark from a teammate's kick. Likewise, a tackle in general play is treated differently to tackles at stoppages. The zone of events was also used, with three possible options:

1. Defensive 50
2. Midfield
3. Forward 50

As an example of how a player's involvement is then represented by the above variables, we consider Fremantle's Nat Fyfe from 2015. His most common involvement was via a handball receive in the midfield in general play, followed by a gather from hitout in the midfield at stoppage. Table 2 contains his five most-common involvements in games, of the 175 observed combinations across all players in 2015.

| Event Type | Zone | Game State | Count | % Total |
|---|---|---|---|---|
| Handball Receive | Midfield | General Play | 113 | 8.2% |
| Gather from Hitout | Midfield | Stoppage | 93 | 6.8% |
| Inside 50 | Midfield | General Play | 87 | 6.4% |
| Centre Bounce Clearance | Midfield | Stoppage | 70 | 5.1% |
| Hard Ball Get | Midfield | Stoppage | 54 | 3.9% |
| Throw In Clearance | Midfield | Stoppage | 50 | 3.6% |
| Tackle | Midfield | Stoppage | 43 | 3.1% |
| Ineffective Kick | Midfield | Stoppage | 43 | 3.1% |

Table 2: Nat Fyfe's most-common involvements in 2015.

All of a player's involvements are represented as a 175-dimension vector, $W$. Each element of $W$ corresponds to a unique combination of event type, zone and game state, and its magnitude is the percentage of

All of a player's involvements are represented as a 175-dimension vector, $\omega$. Each element of $\omega$ corresponds to a unique combination of event type, zone and game state, and its magnitude is the percentage of the player's total involvements for that particular combination. The similarity between two players ($i$ and $j$) is then calculated as a linear transformation of the vector angle between their individual involvements vectors.

$$\theta(i,j) = 100\% - \frac{2}{\pi}\arccos\left(\frac{\omega_i \bullet \omega_j}{|\omega_i| \times |\omega_j|}\right) \tag{1}$$

From (1), players with a vector angle of $\pi/2$ are completely orthogonal, so have a similarity of 0%. Those with a vector angle approaching zero radians are considered 100% similar.

## 3. RESULTS

### PLAYER SIMILARITY

The two most-similar players from the 2015 AFL season were West Coast's Josh J. Kennedy and GWS Giants' Jeremy Cameron. The 10 most similar combinations from the 2015 season can be seen below in Table 3. It is interesting to note that the ninth-highest pair based on similarity is Andrew Gaff and Chris Masten – teammates at West Coast who both play on the wing. Also interesting to note is a small ring of players who are similar to each other – Jake Lloyd, Sam Gibson, Steele Sidebottom and Brandon Ellis. All players are also wingmen who tend to play behind the ball in a defensive position. This ring of players hints that similarity can be used as a means of clustering players.

| Player 1 | Club | Player 2 | Club | Similarity |
|---|---|---|---|---|
| Josh J. Kennedy | West Coast Eagles | Jeremy Cameron | GWS Giants | 92% |
| Ivan Maric | Richmond | Zac Smith | Gold Coast Suns | 89% |
| Jack Macrae | Western Bulldogs | Jack Steven | St Kilda | 89% |
| Jake Lloyd | Sydney Swans | Sam Gibson | North Melbourne | 89% |
| Steele Sidebottom | Collingwood | Sam Gibson | North Melbourne | 89% |
| Brandon Ellis | Richmond | Jake Lloyd | Sydney Swans | 89% |
| Brandon Ellis | Richmond | Sam Gibson | North Melbourne | 89% |
| Billy Longer | St Kilda | Todd Goldstein | North Melbourne | 88% |
| Andrew Gaff | West Coast Eagles | Chris Masten | West Coast Eagles | 88% |
| Josh J. Kennedy | West Coast Eagles | Jack Riewoldt | Richmond | 88% |

Table 3: Most-similar AFL players 2015 (minimum 10 games played).

At the other end of the scale, we can look at the most unique players in the AFL – those with low similarity to all of their peers. Table 4 contains the 10 AFL players from 2015 with the lowest similarity to their most-similar player. This list is dominated by players who played multiple roles across the course of the season, such as Jordan Roughead (who split his time between defence and ruck), Mark Blicavs (ruck and midfield), Kurt Tippett (ruck and forward) and Sam Day (defence and forward).

| Player | Club | Most Similar | Club | Similarity |
|---|---|---|---|---|
| Jordan Roughead | Western Bulldogs | Ben McEvoy | Hawthorn | 69% |
| Adam Saad | Gold Coast Suns | Kade Kolodjashnij | Gold Coast Suns | 70% |
| Mark Blicavs | Geelong Cats | Matthew Kreuzer | Carlton | 70% |
| Kurt Tippett | Sydney Swans | Zac Clarke | Fremantle | 70% |
| Jeremy Howe | Melbourne | Jamie Macmillan | North Melbourne | 72% |
| Patrick Ambrose | Essendon | Isaac Heeney | Sydney Swans | 72% |
| Jake Neade | Port Adelaide | Robin Nahas | North Melbourne | 73% |
| Steven Morris | Richmond | Tendai Mzungu | Fremantle | 73% |
| Sam Day | Gold Coast Suns | Jake Carlisle | Essendon | 73% |
| Fletcher Roberts | Western Bulldogs | Steven May | Gold Coast Suns | 74% |

Table 4: Players least similar to their most-similar players.

We can also use this similarity measure to compare upcoming junior players to their AFL counterparts. This can be used by recruiting staff to better identify player strengths and weaknesses when scouting players, and by the media and general public to gain a better understanding of a player's skill set and potential role in games. Table 5 contains the top-5 draft picks from the 2015 National Draft, with their three most-similar AFL equivalents.

| Player | Club | Player 1 | Player 2 | Player 3 |
|---|---|---|---|---|
| Jacob Weitering | Carlton | Michael Hurley | Dane Rampe | Michael Firrito |
| Josh Schache | Brisbane Lions | Jack Redpath | Jesse White | Eddie Betts |
| Callum Mills | Sydney Swans | Lewis Taylor | Tom Cutler | Rhyce Shaw |
| Clayton Oliver | Melbourne | Josh P. Kennedy | Ben Newton | Mitch Wallis |
| Darcy Parish | Essendon | Dayne Beams | Josh P. Kennedy | Trent Cotchin |

Table 5: Similar players to 2015 Top-5 draft picks.

Likewise, we can perform the same comparison for female players. A national women's competition will not be formally introduced until 2017 but exhibition games were played between Melbourne and Western Bulldogs over the last two seasons, with Champion Data recording statistics from these games. Table 6 below contains similar players for a selection of women from these games from 2015.

| Player | Player 1 | Player 2 | Player 3 |
|---|---|---|---|
| Daisy Pearce | Matt Priddis | Tom Mitchell | Jobe Watson |
| Katie Brennan | Danyle Pearce | Paul Seedsman | Stephen Hill |
| Meg Hutchins | Lynden Dunn | Heath Shaw | Tyson Goldsack |
| Melissa Hickey | Justin Clarke | Phil Davis | Daniel Merrett |
| Tayla Harris | Nic Naitanui | Todd Goldstein | Tom Hickey |

Table 6: Similar players to 2015 women's players.

PLAYER EFFICIENCY

By comparing player performance to those similar to them, we can get a better understanding of their value. Some rating systems may contain unintended biases against particular player types, or certain roles may lead to less opportunity for players to gain points in a rating system. For the following measure we use the Official AFL Player Ratings, as outlined in Jackson (2016), to measure player performance. Table 7 contains a list of the 10 best players when comparing their output in the Player Ratings system to the average output of their five most similar players. We define a player's efficiency (EFF) as:

$$\mathrm{EFF}_i = \frac{\overline{x}_i}{\frac{1}{5}\sum_{j=1}^{5}\overline{x}_{S_{i,j}}} - 100\%$$

where $\overline{x}_i$ is the average score per game of player $i$, and $S_{i,j}$ is the $j^{th}$ most-similar player to player $i$.

| Player | Club | Average | Average of Peers | Efficiency |
|---|---|---|---|---|
| Dayne Zorko | Brisbane Lions | 13.1 | 7.4 | +77% |
| Alex Rance | Richmond | 13.8 | 8.0 | +71% |
| Easton Wood | Western Bulldogs | 14.9 | 8.9 | +68% |
| Todd Goldstein | North Melbourne | 17.0 | 10.9 | +55% |
| Steven May | Gold Coast Suns | 11.6 | 7.6 | +53% |

Table 7: Efficiency of players relative to their five most similar players.

Table 8 below shows the players who benefit the most from looking at performance in terms of efficiency rather than as a raw average points per game – calculated as competition rank for raw points divided by competition rank for efficiency. Brisbane's Dayne Zorko is 42nd in the competition for average points per game, but is No.1 for efficiency relative to similar players. Another player with a large difference is Essendon's Mark Baguley – a defender – who moves from 254th for raw points to 21st for efficiency.

| Player | Club | Average | Rank | Peers | Efficiency | Rank |
|---|---|---|---|---|---|---|
| Dayne Zorko | Brisbane Lions | 13.1 | 42 | 7.4 | +77% | 1 |
| Steven May | Gold Coast Suns | 11.6 | 96 | 7.6 | +53% | 5 |
| Tom Lynch | Adelaide Crows | 11.8 | 82 | 7.9 | +50% | 6 |
| Alex Rance | Richmond | 13.8 | 25 | 8.0 | +71% | 2 |
| Mark Baguley | Essendon | 8.6 | 254 | 6.3 | +37% | 21 |
| Corey Enright | Geelong Cats | 10.7 | 134 | 7.4 | +44% | 13 |
| Travis Varcoe | Collingwood | 9.6 | 189 | 6.9 | +40% | 19 |
| Tom Scully | GWS Giants | 10.1 | 164 | 7.2 | +41% | 17 |
| Cale Hooker | Essendon | 10.4 | 151 | 7.3 | +42% | 16 |
| Devon Smith | GWS Giants | 11.9 | 81 | 8.0 | +48% | 9 |

Table 8: Top-10 players for competition rank for efficiency relative to raw points.

NETWORK PLOTS

Plotting all of a club's players using a network plot allows us to visualise the make-up of the club's roster. Figure 1 below shows such a plot for the Western Bulldogs of 2015, though only the players who had 10 or more appearances for the season. Each player is linked to his three most-similar teammates. From the below plot we can quickly identify some hybrid players – those who play across multiple positions or fulfil multiple roles within games.

- Easton Wood acts as a link between the shutdown defenders (Hamling, Talia, Morris and Roberts) and the more attacking defenders (Boyd, Johannisen, Murphy and Biggs).
- Stewart Crameri acts as a link between the permanent forwards (Stringer, Dickson, Boyd and Redpath) and the mid-forwards who rotate between the forward-line and the midfield (Grant, Dale, Hunter, Daniela and Hrovat).

Note that with four players filling the ruck role throughout the year for the Western Bulldogs, they form an isolated network unconnected to other position types.


Figure 1: Network plot for the 2015 Western Bulldogs.

In Figure 2 the same representation is shown for Hawthorn. It is clear that at either end of the network there is a distinct grouping of permanent forwards (Gunston, Roughead, Breust, Schoenmakers, Rioli and Puopolo) and permanent defenders (Frawley, Litherland, Stratton, Gibson and Lake). The two permanent ruckmen used (Ceglar and McEvoy) sit between the club's midfielders and David Hale, who was predominantly a forward, but spent short periods in the ruck.

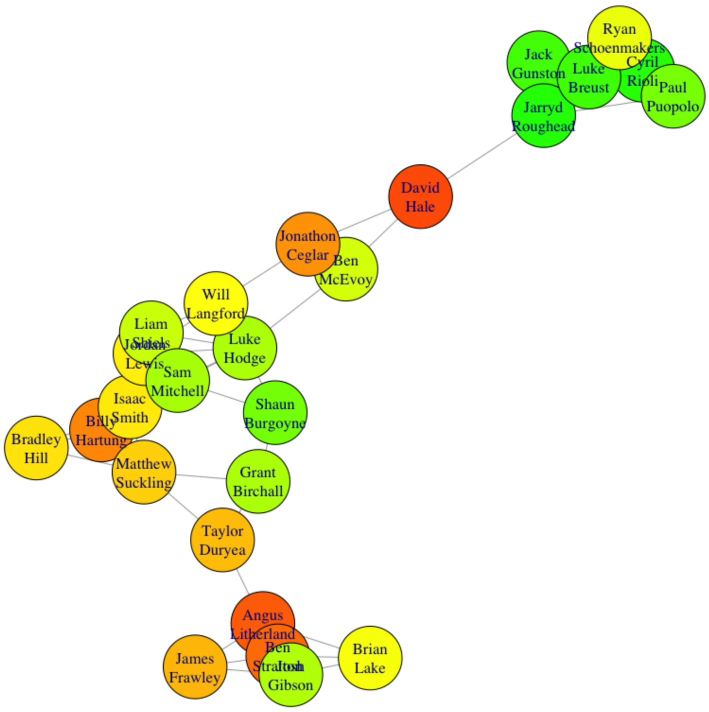



Figure 2: Network plot for the 2015 Western Bulldogs.

## 4. CONCLUSION

It was shown that using vector angles on a breakdown of player involvements produces a viable measure of player similarity. Comparisons across competitions can be used to preview the careers of new players or to gain greater context on the performance of unknown players. Further research is suggested in clustering players into discrete position classes to better enable comparison of player performance.

## References

Jackson, K. (2016). Assessing Player Performance in Australian Football Using Spatial Data. PhD Thesis, Swinburne University of Technology.

Pyne, D., Gardner, A., Sheehan, K. & Hopkins, W. (2006). Fitness testing and career progression in AFL football. Journal of Science and Medicine in Sport, 8, 321-332

Sargent, J. & Bedford, A. (2010). Long-Distance Relationships: Positional Classification of Australian Football League Players. Proceedings of the tenth Australasian conference on Mathematics and Computers in Sport, 97-102.

# FIXTURE DIFFICULTY AND TEAM PERFORMANCE MODELS FOR USE IN THE AUSTRALIAN FOOTBALL LEAGUE

Casey Josman [a,c], Ritu Gupta [a], and Sam Robertson [b]

[a] *Department of Mathematics and Statistics, Curtin University, Perth, Western Australia*
[b] *Victoria University, Institute of Sport, Exercise & Active Living, Melbourne, Australia*
[c] *Corresponding author: Casey.Josman@curtin.edu.au*

## Abstract

Two major talking points amongst both team sports fans and personnel are a) season difficulty and b) relative team performance. Although heavily debated in a variety of forums, it is important that these constructs be quantified in an effort to understand fully the factors which influence fixture difficulty and performance and allow coaches and training staff to make informed training and in-game decisions. Hence, we present methodologies for the quantification of both season difficulty and relative team performance based on data that is entirely available prior to the start of each match. Data was obtained for all matches played during the 2010 – 2015 AFL premiership seasons, and contained information pertaining to the season, round, teams, venue, scores, and team ranks for each match with additional measures of form (head to head and individual past performance ratios) being calculated from the raw data. A multivariate logistic regression (MLR) model for predicting the probability of a win for a team was developed that during validation achieved an accuracy of 67.8%. Based on this MLR model, we present a framework for the quantification of season difficulty and the evaluation of a team's performance throughout a given season. These models make use of rank differentials, Bernoulli simulation, and a newly proposed risk matrix for point assignment respectively. When viewed against an objective analysis of actual team performance these models perform comparatively well and as such can be used to make informed training decisions.

*Keywords*: AFL, Fixture Difficulty, Multivariate Logistic Regression, Performance Analysis

## 1. INTRODUCTION

Although debated across many forums, both the concepts of season difficulty and relative team performance within the Australian Football League (AFL) have yet to be quantified by mathematical means. The literature to date has focussed on match outcome prediction (Maszczyk et al. 2014), optimal betting strategies (Crowder et al. 2002), or rudimentary analysis of key performance indicators (Jones, Mellalieu and James 2004) in sports other than AFL. Currently only simple approximations for season and match difficulty are used within the literature, for instance differentials of ELO style ratings (Hvattum and Arntzen 2010) and simplistic probability based models such as Bradley-Terry type models (McHale and Morton 2011).

The aims of this research were twofold; firstly, to objectively evaluate the difficulty of a given season using mathematical means, and secondly, to develop a method by which it is possible to quantify team performance over a given season. To achieve this, we present a novel methodology for quantifying season difficulty and team performance. Guided by practical knowledge and current literature, we utilise various models based on multivariate logistic regression which make use of rank differentials, Bernoulli simulation, and a new risk matrix for the awarding of points.

## 2. METHODS

DATA

Data for this study was acquired from afltables.com (2015) and consisted of full match data for all games played during the 2010-2015 AFL premiership seasons (1210 matches), more specifically data pertaining to; season, round, home team, away team, home rank, away rank, home score, and away score were utilised. An initial screening of the data revealed that 12 of the matches concluded with a draw and as such were removed due to their infrequent occurrence (0.9917%).

For each retained match both home and away teams were reassigned using a robust methodology which aims to provide an unbiased definition of the home/away assignments for each match. The rationale behind this was to remove bias introduced by inter-club politics and take into account stadium conditions (location and crowd composition) which have been shown to directly affect player performance (Reicher 2001). For any pairing of teams $(i, j) \in (H, A)$ in a match $m$ the home and away teams are defined as follows; if either team $i$ or $j$ are playing at their home ground then assign the home team accordingly, however, if both teams $i$ and $j$ share the same home ground or are both playing an away game then assign the home team to that team which has the highest official membership number. This reassignment yields a home team win percentage of 0.581 which holds with the paradigm of home advantage outlined by Stefani and Clarke (1992) and Clarke (2005) as well as remaining statistically similar to the original data which has a home team win percentage of 0.584 ($\Delta$<0.005).

TEAM BASED STATISTICS CALCULATION

In addition to the data above, various other metrics with respect to team performance and form were calculated. For a given match $m$ between home and away teams $i$ and $j$, the result $R_{i,j,m} = 1$ if team $i$ wins and $R_{i,j,m} = 0$ if team $i$ loses. Furthermore, let $R_{i,j,m} = 0$ if the m^th^ match was not between teams $i$ and $j$. Therefore, for a set of matches played between a home team $H$ and an away team $A$, team based statistics were calculated as follows;

- Head2Head: The percentage of games over the past 5 games for which the home team has won against the away team.

$$\frac{\sum_{g=m-6}^{m-1}\left(R_{H,A,g}\right)}{5} \tag{1}$$

- PastHome: The percentage of games over the past 4 games for which the home team has won against any opponent.

$$\frac{\sum_{j=1}^{4}\sum_{g=m-5}^{m-1}\left(R_{H,j,g}\right)}{4} \tag{2}$$

- PastAway: The percentage of games over the past 4 games for which the away team has won against any opponent.

$$\frac{\sum_{i=1}^{4}\sum_{g=m-5}^{m-1}\left(R_{A,i,g}\right)}{4} \tag{3}$$

PREDICTIVE ALGORITHM

Throughout this study both result and win probability prediction were performed using multivariate logistic regression (MLR) and made use of the features $F$ described above. MLR is a generalised linear model that has been successfully used for ex-ante prediction (Lopez and Matthews 2014) and in our application achieves an accuracy of 67.8%. Let $C(F)$ and $\mathbb{C}(F)$ represent predictions of match result and win probability (with respect to the home team) respectively, with the MLR model being specified as

$$\ln\left[\frac{\mathbb{C}(F)}{1-\mathbb{C}(F)}\right] = \beta F \tag{4}$$

where $\beta$ are the unknown coefficients to be estimated, $C(F)$ is derived from a logistic regression fit function which classifies a win as follows:

$$R_{H,A,m} = \begin{cases} 1 & \mathbb{C}(F) \geq c \\ 0 & otherwise \end{cases} \tag{5}$$

where $c$ is the cut-off point for the logistic regression fit function which is taken as the intersection of both sensitivity $\Pr(\widehat{Y}_i = 1\,|Y_i = 1)$ and specificity $\Pr(\widehat{Y}_i = 0\,|Y_i = 0)$, and $\mathbb{C}(F)$ is taken as the direct output of the MLR.

SEASON DIFFICULTY

The difficulty $D_{\mathbb{T},\mathcal{R}}$ of a season for a given team $\mathbb{T}$ starting the season at rank $\mathcal{R}$ can be defined using one of two models; the previous season ranking model (PSR) which is a simple linear style model, and the season ranking simulation (SRS) model which is a model predicated on the basis of our MLR model.

Model 1: PSR

The difficulty $D_{\mathbb{T},\mathcal{R}}$ derived from the PSR model is defined as the sum of the differences in ranking between the reference team and their opponents during their 11 home and 11 away games ($hg$ and $ag$ respectively) during a given season.

$$D_{\mathbb{T},\mathcal{R}} = \sum_{hg=1}^{11}\left(\mathcal{R}_{\mathbb{T},hg} - \mathcal{R}_{A,hg}\right) + \sum_{ag=1}^{11}\left(\mathcal{R}_{\mathbb{T},ag} - \mathcal{R}_{H,ag}\right) \tag{6}$$

Scores are then approximated as standardised random variables as per (7) by setting both mean and standard deviation as the arithmetic mean and range of $\mathbb{A}_{\mathbb{T},\mathcal{R}}$ and $\mathbb{B}_{\mathbb{T},\mathcal{R}}$ respectively, where $\mathbb{A}_{\mathbb{T},\mathcal{R}}$ and $\mathbb{B}_{\mathbb{T},\mathcal{R}}$ are the minimum and maximum possible difficulty ratings for a given team and starting rank (as outlined by the AFL Commission) respectively, with values less than 0 indicating an easier than average season and vice versa.

$$D^*_{\mathbb{T},\mathcal{R}} = \frac{D_{\mathbb{T},\mathcal{R}} - \mu_{\mathbb{T},\mathcal{R}}}{\sigma_{\mathbb{T},\mathcal{R}}}, where\ \mu_{\mathbb{T},\mathcal{R}} = \frac{\mathbb{A}_{\mathbb{T},\mathcal{R}} - \mathbb{B}_{\mathbb{T},\mathcal{R}}}{2}\ and\ \sigma_{\mathbb{T},\mathcal{R}} = \mathbb{B}_{\mathbb{T},\mathcal{R}} - \mathbb{A}_{\mathbb{T},\mathcal{R}} \tag{7}$$

The AFL Commission (Australian Football League 2015) have outlined the following guidelines for the setting of fixtures (accurate as of the 2015 AFL season); each team is to play 22 games over a period of 25 weeks with each team playing each other team at least once. Teams ranked 1-6 at the beginning of the season will then play either 2 or 3 additional games against other teams ranked 1-6, either 1 or 2 additional games against teams ranked 7 to 12, or either 0 or 1 additional games against teams ranked 13-18. Teams ranked 7-12 at the beginning of the season will then play either 1 or 2 additional games against teams ranked 1-6, either 2 or 3 additional games against other teams ranked 7 to 12, or either 1 or 2 additional games against teams ranked 13-18. Teams ranked 13-18 at the beginning of the season will then play either 0 or 1 additional games against teams ranked 1-6, either 1 or 2 additional games against teams ranked 7 to 12, or either 2 or 3 additional games against other teams ranked 13-18. From the above guidelines it is possible to generate a list of maximum ($\mathbb{B}_{\mathbb{T},\mathcal{R}}$) and minimum ($\mathbb{A}_{\mathbb{T},\mathcal{R}}$) difficulty rating values for each team given their starting rank and number of scheduled games $G^{min}_{\mathbb{T},j}$ and $G^{max}_{\mathbb{T},j}$ against team $j$, where $G^{min}_{\mathbb{T},j}$ and $G^{max}_{\mathbb{T},j}$ are the easiest and hardest sets of scheduled games respectively.

$$\mathbb{A}_{\mathbb{T},\mathcal{R}} = 22\mathcal{R}_{\mathbb{T}} - \sum_{\substack{j=1 \\ j\neq\mathbb{T}}}^{18} \mathcal{R}_j\, G^{min}_{\mathbb{T},j} \tag{8}$$

$$\mathbb{B}_{\mathbb{T},\mathcal{R}} = 22\mathcal{R}_{\mathbb{T}} - \sum_{\substack{j=1 \\ j\neq\mathbb{T}}}^{18} \mathcal{R}_j\, G^{max}_{\mathbb{T},j} \tag{9}$$

Model 2: SRS

The SRS model is a hybrid simulation model combining aspects of result prediction, Bernoulli simulation, and linear modelling. Using this model the difficulty $D_{\mathbb{T},\mathcal{R}}$ is derived as the difference between a team's rankings $\mathcal{R}_{\mathbb{T},y}$ at the end of the current and previous seasons

$$D_{\mathbb{T},\mathcal{R}} = \mathcal{R}_{\mathbb{T},y} - \mathcal{R}_{\mathbb{T},y-1} \tag{10}$$

where a team's ranking at the end of the current season is obtained by simulating the season's fixture for possible match points based on the MLR model.

In this simulation as per a normal AFL season, a team is awarded 4 points for each win with the total number of points being averaged over all simulations and end of season rankings $\mathcal{R}_{\mathbb{T},y}$ assigned accordingly. Differences are then calculated as per (10) with a negative difference indicating an easier season and vice versa.

TEAM PERFORMANCE

The current performance $\mathbb{P}_{\mathbb{T},y}$ for a team $\mathbb{T}$ during season $y$ is defined by the static penalty (SP) model as

$$\mathbb{P}_{\mathbb{T},y} = \sum_{m=1}^{22} \mathcal{P}_{\mathbb{T},m,y} \tag{11}$$

where $\mathcal{P}_{\mathbb{T},m,y}$ is the point value awarded to team $\mathbb{T}$ after match $m$ during season $y$. For a given win probability $\mathbb{C}(F)$ corresponding to a reference team $\mathbb{T}$ a simple method for quantifying match performance is defined as

$$\mathcal{P}_{\mathbb{T},m,y} = \begin{cases} \min\left(25, \dfrac{1}{\mathbb{C}(F)}\right) & Score_{\mathbb{T}} > Score_A \\ \max\left(-25, -\dfrac{1}{1-\mathbb{C}(F)}\right) & otherwise \end{cases} \tag{12}$$

with larger values indicating better team performance.

**3. RESULTS**

In summary the MLR model used throughout this research achieved a validated accuracy of 67.8%, has a residual deviance of 749.6 with 693 degrees of freedom, $\chi^2$=638.09, and a p-value=0.06712 which at a 5% level of significance would indicate no evidence of a lack of fit. However, as the model is not saturated a more appropriate test would be the Hosmer-Lemeshow test (Hosmer Jr and Lemeshow 2004) which has $\chi^2$=5.48, and a p-value=0.7 which at a 5% level of significance indicates a significantly good fit.

A cursory look at the win probabilities generated by the MLR model (Figure 1) would indicate that teams such as St Kilda and Melbourne have the hardest season and teams such as Geelong and Sydney have the easiest season. However, as is the nature of a competitive game such as AFL the team with the easiest season does not necessarily perform the best.

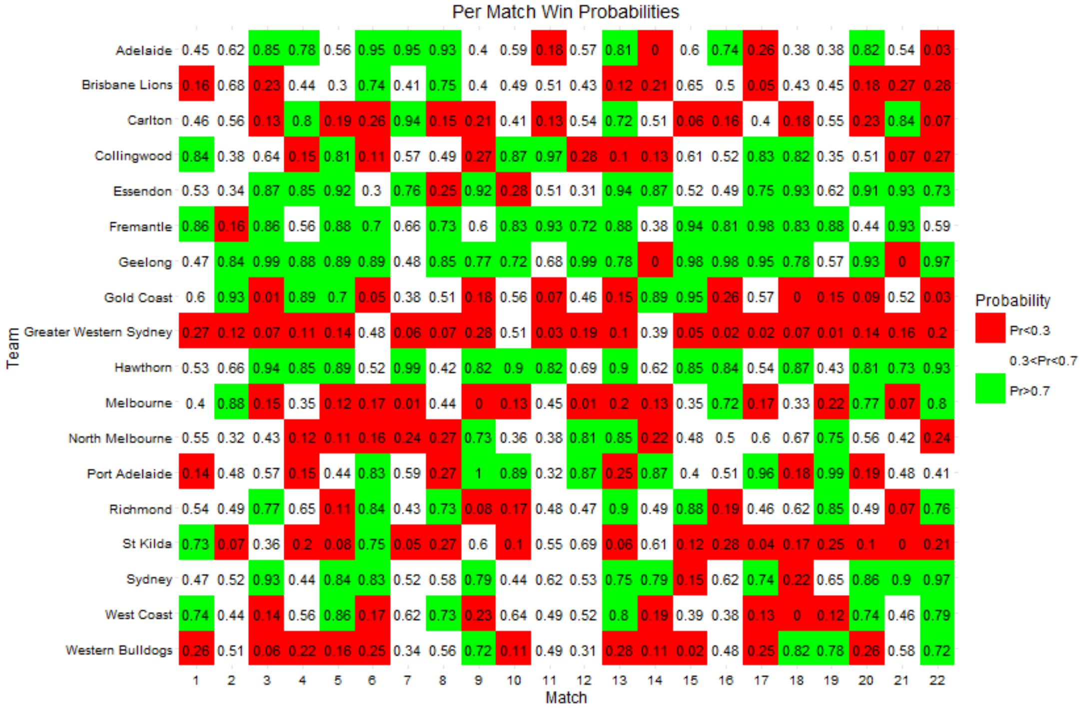



Figure 1: Per Match Win Probabilities for the 2015 AFL Season

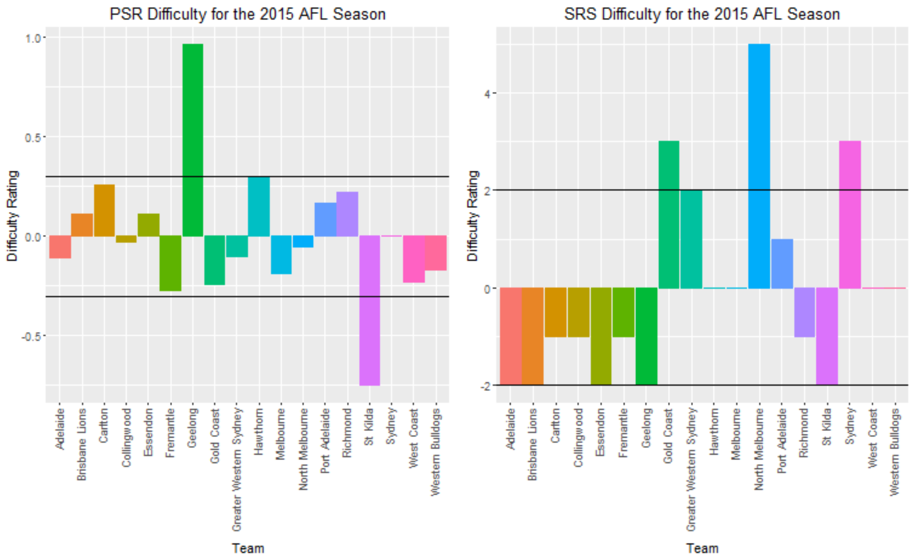
## SEASON DIFFICULTY

Figure 2 presents the season difficulty (PSR and SRS respectively) for each team during the 2015 AFL premiership season, contrary to our cursory analysis the PRS model predicts St Kilda and Geelong to have the easiest and hardest season respectively. However, after a closer look we can see that the difficulty ratings for St Kilda and Geelong are clearly outliers and can be attributed to the simplistic nature of the model. Another observation that can be made is that the remaining 16 teams have a difficulty rating between -0.3 and 0.3 and as such can be said to have a relatively fair season.

Using the results generated from the SRS model we can see that most of the results lie within the range of -2 to 2 and can therefore be concluded that the season is relatively easy for all teams other than Gold Coast, Sydney, and North Melbourne. Gold Coast and Sydney having only a marginally more difficult season with difficulty scores of 3 and North Melbourne having a significantly more difficult season with a difficulty score of 5.

The variance in results amongst the two models can be seen as either a result of the increasing complexity of each model or the inherent differences in each model's design. The PSR model is based on standardized rank differentials and as such aims to negate the unbalanced structure of the AFL season, whereas the SRS model aims to simulate the outcome of the AFL season. Surprisingly though, the PSR model in its simplicity shares similar characteristics with the preseason model employed by Champion Data which makes use of points differentials from the 2014 season which provided teams maintain a relatively stable level of performance can be seen as a proxy for the final rankings of the 2014 season which are employed by the PSR model.

The SRS model similarly predicts a reasonably easy season for most of the teams, however, it singles out Gold Coast, North Melbourne, and Sydney as having the hardest season whereas the Champion Data model rates their seasons as moderately difficult. This is more than likely a compounding effect realized through the simulation of an already biased fixture.




(a) PSR Difficulty (b) SRS Difficulty

Figure 2: Season Difficulty Results (Difficulties within horizontal boundaries represent relatively easy fixtures)

TEAM PERFORMANCE

Figure 2 presents the results from the SP model. Once again Geelong has appeared at the bottom of our list and St Kilda towards to top, and so it could be concluded that the previous season ranking model and the variable penalty difficulty model complement each other in their simplicity. We can also conclude that there are three performance clusters; with Gold Coast, Essendon, and Geelong performing the worst, Adelaide, Brisbane Lions, Carlton, and Collingwood performing adequately, and the other teams performing very well.

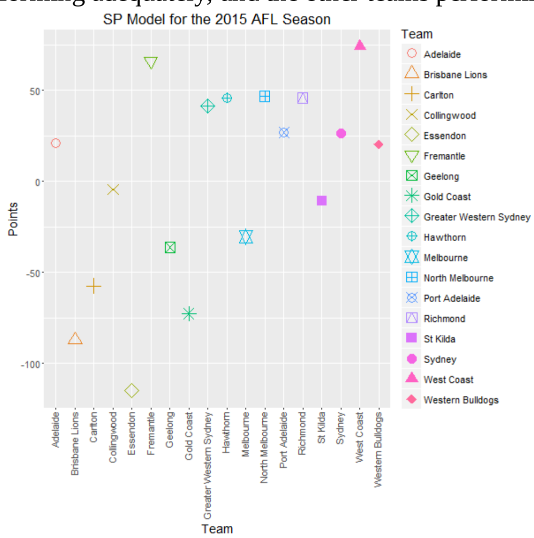



Figure 3: Static Penalty Model Performance Plot

## 4. DISCUSSION

The aim of this study was to determine whether it is possible to mathematically quantify both season difficulty and team performance. With respect to the MLR model's accuracy, it achieves similar results to those in the literature. Baker and McHale achieved accuracies of 63.6 and 66.9% respectively using a continuous-time Markov process to predict the outcomes of National Football League (NFL) games, Akhtar and Scarf (2012) achieved a 59.6% accuracy for predicting the ex-ante outcomes of cricket matches when using a MLR model, and Carbone, Corke, and Moisiadis (2016) achieved accuracies of 63 and 55.7% respectively using an ELO based method for predicting National Rugby League (NRL) match outcomes. Our results compare favourably to those in the literature. However, it would be possible to use other features which are not routinely available prior to the beginning of the season (number of best players available, injuries, etc.) to increase model efficacy.

Whilst the predictive accuracy of our model compares similarly to those in the literature – all of these models assume independence between matches. However, it can be safe to say that match results are subject to some form of dependence. Nevertheless, violation of the independence assumption does not significantly impact the final results due to the scale of our data (Heo and Leon 2005).

The season difficulty models were initially designed with model simplicity in mind (PSR) and then graduated to a more complex simulation model (SRS), the rationale behind the PSR model is that it provides a model based on the most simplistic (and in this case most telling) metric of opponent difficulty (previous season ranking), while the SRS model attempts to simulate the outcome of a given season and then draw inferences with respect to relative fixture difficulty.

The team performance model (SP) was designed using a truncated risk matrix such that the points assigned to a team who wins a very easy match (Pr(Win)>0.7) are significantly smaller in magnitude than points assigned to a team who wins a very hard match (Pr(Win)<0.3) with the inverse true for a team who loses a match. The rationale behind this design is that it is believed to be able to more accurately capture the real world implications of winning and losing matches of varying difficulty. The significantly larger negative results from the SP model are due to the heavy weighting assigned for winning and losing hard and easy games respectively. The coefficients and parameters of the risk matrix can also be altered in accordance to the MLR model and coaching decisions. Hence, the methodology presented in this paper can be utilised for other competitive team sports.

## 5. CONCLUSIONS

An important characteristic to be aware of when creating models for the AFL is that unlike a majority of European sports the AFL has an inherently biased fixture structure which is dictated by inter-club politics. With this in mind it is more than feasible to suggest that teams and coaches not only be judged by their wins and losses in a given season but also by their performance relative to the difficulty of their respective fixtures. Our model may also be repurposed for use in an applied sense. Firstly, future model features relating to relative team performance could be derived and may consist of data such as a team’s performance over or under a given fixtures difficulty, and secondly, our results may be used to assist with fixture planning and fairness evaluation by the AFL.

**Acknowledgements**

We wish to thank Paul from afltables.com for supplying us with data for this study.

**References**

AFL Tables. 2015. *Afl-Vfl Match, Player and Coaching Stats, Records and Lists*. http://afltables.com/afl/afl_index.html.

Akhtar, Sohail, and Philip Scarf. 2012. "Forecasting Test Cricket Match Outcomes in Play." *International Journal of Forecasting* 28 (3): 632-643. doi: http://dx.doi.org/10.1016/j.ijforecast.2011.08.005.

Australian Football League. 2015. Laws of Australian Football 2015. Melbourne: Australian Football League.

Carbone, Joel, Tony Corke, and Frank Moisiadis. 2016. "The Rugby League Prediction Model: Using an Elo-Based Approach to Predict the Outcome of National Rugby League (Nrl) Matches." *International Educational Scientific Research Journal* 2 (5).

Clarke, Stephen R. 2005. "Home Advantage in the Australian Football League." *Journal of Sports Sciences* 23 (4): 375-385.

Crowder, Martin, Mark Dixon, Anthony Ledford, and Mike Robinson. 2002. "Dynamic Modelling and Prediction of English Football League Matches for Betting." *Journal of the Royal Statistical Society: Series D (The Statistician)* 51 (2): 157-168.

Heo, Moonseong, and Andrew C Leon. 2005. "Comparison of Statistical Methods for Analysis of Clustered Binary Observations." *Statistics in medicine* 24 (6): 911-923.

Hosmer Jr, David W, and Stanley Lemeshow. 2004. *Applied Logistic Regression*: John Wiley & Sons.

Hvattum, Lars Magnus, and Halvard Arntzen. 2010. "Using Elo Ratings for Match Result Prediction in Association Football." *International Journal of forecasting* 26 (3): 460-470.

Jones, Nicholas MP, Stephen D Mellalieu, and Nic James. 2004. "Team Performance Indicators as a Function of Winning and Losing in Rugby Union." *International Journal of Performance Analysis in Sport* 4 (1): 61-71.

Lopez, Michael J, and Gregory Matthews. 2014. "Building an Ncaa Mens Basketball Predictive Model and Quantifying Its Success." *arXiv preprint arXiv:1412.0248*.

Maszczyk, Adam, Artur Gołaś, Przemysław Pietraszewski, Robert Roczniok, Adam Zając, and Arkadiusz Stanula. 2014. "Application of Neural and Regression Models in Sports Results Prediction." *Procedia - Social and Behavioral Sciences* 117 (0): 482-487. doi: http://dx.doi.org/10.1016/j.sbspro.2014.02.249.

McHale, Ian, and Alex Morton. 2011. "A Bradley-Terry Type Model for Forecasting Tennis Match Results." *International Journal of Forecasting* 27 (2): 619-630. doi: http://dx.doi.org/10.1016/j.ijforecast.2010.04.004.

Reicher, Stephen. 2001. "The Psychology of Crowd Dynamics." *Blackwell handbook of social psychology: Group processes*: 182-208.

Stefani, Raymond, and Stephen Clarke. 1992. "Predictions and Home Advantage for Australian Rules Football." *Journal of Applied Statistics* 19 (2): 251-261.

# IDENTIFICATION AND MEASUREMENT OF LUCK IN SPORT

Darren O'Shaughnessy [a,b,c]

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

## Abstract

In Australian Rules Football, play is started each quarter, and after each goal, by the umpire bouncing the ball in the centre of the ground to be contested by two ruckmen while several shorter players rove the vicinity. There are about 28 of these events each match, and it is common wisdom that winning clear possession here is a Key Performance Indicator. When a team wins the ball from this restart, it immediately enters its attacking zone 70% of the time, and scores directly from 21% of its clearances without an opponent touching the ball.
Whilst evaluating Hawthorn Football Club's successful 2014 season, I discovered three remarkable disruptions to this common sense interpretation:

1. Despite full-time stoppage coaches and massive investment in strategies, the win-loss ratio of clubs in an entire season's data was essentially a coin flip. Only one club won centre clearances at a rate outside ±1.5 standard errors in the mean compared with a Bernoulli trial model, and the standard deviation of the 18 clubs' SEMs was 1.06.
2. The centre clearance statistic had zero reliability; in fact teams were slightly more likely to get a negative differential in the match following a positive one.
3. Over the season, there was a strong negative correlation ($r = -0.68$) between clubs' clearance win rate and the conversion of those opportunities into scoreboard equity (O'Shaughnessy, 2006). This suggests teams that invest extra resources in winning the "coin-flip" have depleted resources at more vital locations.

Hawthorn FC embarked on a *gedankenexperiment* where the coaches were asked to consider how they would structure their resources if, instead of competing to win the clearance, the umpire simply flipped a coin and gave the ball to the winner.
This paper explores the implications of classifying some events as mostly luck – a spontaneous breaking of symmetry – and how a sports team or player might measure performance accounting for known sources of variation.

**Keywords**: Luck, performance analysis, Australian Rules Football, AFL

## 1. INTRODUCTION

As the field of performance analysis matures, identification of key performance indicators in sporting contests has expanded. While some new metrics – such as Expected Goals in soccer – are explicitly attempting to reduce random variation, coaches and analysts have often been reluctant to re-evaluate the meaning of traditional KPIs that have always been part of their reporting and coaching methods.

In reality, all actions that are counted on the field of play are affected to some extent by stochastic processes, from the bounce of the ball to the workings of the neuromuscular system (Mauboussin, 2012) to the intervention of officials. The state of play prior to each action provides significant context to the action's outcome, but is itself affected by both systematic and random factors.

Additionally, statistical models that attempt to predict team success based on combinations of these KPIs are affected by several confounding factors including

- Small sample sizes of perhaps a couple of dozen games per year
- Opposition effects: each contest is a dynamic exchange (e.g., Gréhaigne & Godbout, 2014) where the opponent has as much effect on the collection of outcomes as the team. This is unlike most other fields of statistical analysis
- Multi-collinearity and interconnectedness between indicators

We would to identify those actions which are indicative of skill, and those we can classify as luck. Naturally this is a continuum and the proportion of luck in a mixed indicator tends to decrease with the square-root of the number of data points. We would also like to measure the effect of these on the scoreboard. This paper offers some insights into the process in one sport.

## 2. METHODS

THE CENTRE CLEARANCE

One distinctive aspect of Australian Rules Football (colloquially called AFL) is the amount of umpire intervention. There are three field umpires, two boundary umpires, two goal umpires, video reviewers, supervisors and assistants. This befits the somewhat anarchic play with 36 players on the field, no offside rule, full contact, and dozens of player substitutions each quarter. At the beginning of each quarter, and after each goal, an umpire takes possession of the ball in the centre of the oval and either bounces it or throws it into the air[1]. This is called the Centre Bounce, one of three types of *stoppage* where the umpire impels the ball back into play.

At the Centre Bounce, only one designated *ruckman* from each team is able to contest the ball in the air. The ruckman who hits the ball (gaining a *hitout* statistic) attempts to direct it to one of his *rovers*. Three rovers from each team are allowed within the area, and other players arrive from 25m away within five seconds of the bounce. The rover with *first possession* either attempts to clear the ball from the congested area himself with a handball or kick, or passes to a teammate who seeks to achieve this. The first player who effectively passes the ball to a teammate in sufficient space, or successfully clears the centre area without disadvantaging his team, is awarded a *centre clearance* or *centre break*. If neither team can release the ball due to the carrier being tackled, the umpire directs a *secondary stoppage* or *ball-up*.

Figure 1 is a Sankey diagram showing the average flow of ball possession from the Centre Bounce. Where numbers do not sum to 100%, there is leakage due to the play ending without clearance. There are numerous opportunities for individual skill and team strategy to influence the play, from the jump and tap of the ruckman to subtly blocking opponents, commanding space, sharking the ball from the opposition ruck, stealing it at ground level or gaining a free kick. While the best ruckmen can achieve hitout rates of over 65% long-term, *hitouts to advantage* where a teammate has immediate space to make a decision are rarer – about 28% of all CB hitouts are classified this way (not shown in Figure 1, but comprising 56% of the 50% Gathered).

Teams invest considerable resources into stoppages, with professional teams often employing a full-time stoppage coach, engaging ruck consultants and developing playbooks of formations and strategies intended to give the team a significant advantage from these neutral restarts.

For the Centre Clearance analysis in the next section, this paper will consider only those 86% of Centre Bounces that assigned a clearance to a team. Notations from every 2014 match were used ($n$ = 4915 from 207 matches). 70% of *possession chains* from clearances are taken into the attacking zone within 50m of goal without an opponent touching the ball, and of those 30% lead directly to a score.

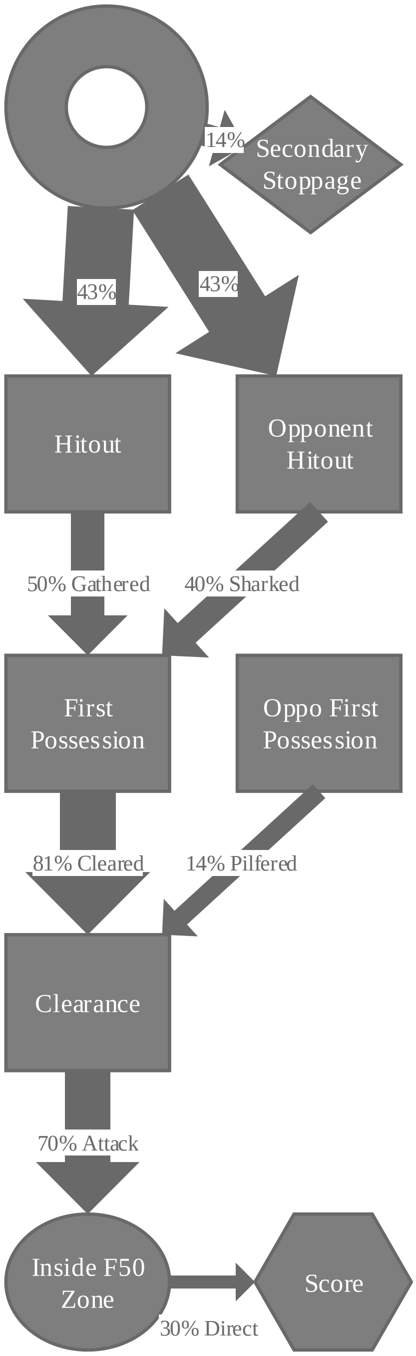



Figure 1: Anatomy of a Centre Clearance

[1] The ball is thrown up when conditions are unfavourable for bouncing, or if the umpire has failed to bounce the ball satisfactorily; these differences are not discussed in this paper

SCOREBOARD EQUITY

O'Shaughnessy (2006) uses the term scoreboard equity to describe the net value of the current state of the ball, with respect to a team's chances of scoring next. Given sufficient time, the Markov process that describes ball movement will be absorbed as a score, worth either +6, +1, -1, or -6 to the team. Scoreboard equity is the expected mean value of the ensemble of absorbing states, weighted by the probability of that score occurring next. Each AFL play can be regarded as a mini-game where teams attempt to optimise equity, in order to maximise their probability of winning the match through realised scores.

Simple data analysis shows that the average equity of a centre clearance in recent years has been +1.06; if the opponent clears the ball the team has equity -1.06. This implies a swing of 2.12 scoreboard points depending on which team wins use of the ball from the centre bounce.

For this paper, the scoreboard equity has been evaluated at known standard states, calculated via parametrisation of empirical data over the seasons 2012-2014. For the purposes of the model, we can regard these states as known intermediate equity values at the location $\boldsymbol{x}$: ball-up by umpire, throw-in by boundary umpire, set shot by either team, running shot by either team. Unlike the 2006 paper where play was tracked until the next score, the model used here truncates the equity calculation to the next stoppage or shot at goal. This should reduce the noise and team-specific effects due to tracking further play after the standard state.

BACKGAMMON ANALYTICS

The game of backgammon, from which the equity theory was adapted, offers reliable ways of analysing the mix of skill and luck employed by both players. The first and most direct method is to compare the player's choice of move with a perfect evaluation, and report the error rate. For instance, if a player has to make 250 decisions during a match, and makes 22 errors with a total loss of equity of 1.4 points versus perfect play, his error rate is reported as -5.6 points per 1000 moves. If his opponent's error rate is -7.3, he can be regarded as playing better whether he won or lost. This measure is not perfect, as it does not take into account the relative difficulty of the decisions, only the extent to which match-winning probability was diminished by each choice.

The second method employed by backgammon analysts is to measure the luck dealt by the dice, then subtract it from the result of the match to leave a residual skill effect. The algorithm evaluates the equity of each of the 21 possible dice rolls, and collects the difference between the mean of those 21 possibilities and the equity of the actual dice roll. This luck contribution can also be scaled using the effect on match-winning probability. In simplified terms, if a player won 11-8 with total luck of +2.6 points, while his opponent's dice produced total luck of -1.4, then the residual displayed skill suggests that he should have lost by one point instead of winning by three.

With a perfect evaluator, these two methods of skill measurement should report the same conclusion. Backgammon software is now superior to human experts, but is imperfect given the size of the decision tree. Nonetheless the two measures are very similar in most real-world situations, validating the use of luck measurement to expose differences in skill.

It is well known among backgammon experts that in a typical match lasting an hour, luck dominates skill. Between opponents of similar expertise, the match is far more likely to be decided in favour of the luckier player than the more skilful one. Since the introduction of sophisticated backgammon software in the past two decades, this fact has become part of the common knowledge and the culture. In contrast in professional sport, there is very little evidence that coaches or commentators have understood this basic law of statistics: if the competitors are close in skill, luck will usually determine the winner of the contest. This applies at the level of player versus player competing for a loose ball, and at the level of team versus team over a match.

In sport there is no such thing as a perfect evaluator, even in principle as players do not take turns using their own random activity generators. However, in the following sections some obvious sources of on-field luck will be partitioned and measured on an equity scale, reducing the error in the measurement of skill difference, compared to the scoreboard.

## 3. RESULTS

CENTRE CLEARANCES

Winning the majority of centre clearances is clearly a factor in winning a match of AFL. As mentioned in the previous section, the effect of the average centre break is an immediate boost of +1.06 points of equity, with the winner scoring next 65% of the time.

The result of a quarter-by-quarter linear regression of score margin vs centre clearance differential yields a slope of 1.00±0.17, while a match-by-match regression of the same data shows a slope of 1.28±0.48. This indicates that the advantage from each clearance persists through the match, but does not show any extra effect that might be due to collinearity between overall team skill and an ability to win centre clearances.

The most startling result came in the analysis of variation in team clearance percentages through the entire 2014 season, as shown in Table 1 and Figure 2. Comparing with a Bernoulli trial model where each team has 50% probability of winning the clearance, 17 of 18 clubs were within ±1.5 standard errors in the mean. The standard deviation of that collection of SEMs is just 1.06, suggesting that the entire season's variation from 50/50 very closely resembles noise.

To test the reliability of the Centre Clearance indicator week to week, each team's centre clearance differential was regressed against its differential in its next match. The correlation coefficient was approximately 0.015, even after adjusting for any home ground advantage effect (home teams averaged 12.4 clearances to 12.0 for the away team). In other words, teams did not seem to be able to maintain any advantage over time, nor did they correct their deficits any more than regression to the mean would suggest.

| Club | Centre Clearances | | | | Equity from Centre Clr | | | |
|---|---|---|---|---|---|---|---|---|
| | Won | Lost | % | SE in Mean | Won | Lost | Diff | Final Position |
| Essendon | 292 | 228 | 56.2% | +2.81 | 1.05 | -0.71 | 1.77 | 7 |
| Adelaide | 311 | 280 | 52.6% | +1.28 | 1.17 | -0.76 | 1.94 | 10 |
| Sydney | 274 | 252 | 52.1% | +0.96 | 1.23 | -0.81 | 2.04 | 2 |
| Fremantle | 274 | 253 | 52.0% | +0.91 | 1.02 | -1.25 | 2.28 | 5 |
| WC Eagles | 278 | 260 | 51.7% | +0.78 | 1.02 | -0.96 | 1.98 | 9 |
| W Bulldogs | 280 | 263 | 51.6% | +0.73 | 0.93 | -1.09 | 2.02 | 14 |
| Carlton | 279 | 273 | 50.5% | +0.26 | 0.93 | -1.06 | 1.99 | 13 |
| Port Adel | 288 | 286 | 50.2% | +0.08 | 1.42 | -0.84 | 2.26 | 3 |
| North Melb | 278 | 280 | 49.8% | -0.08 | 1.12 | -1.04 | 2.16 | 4 |
| St Kilda | 269 | 277 | 49.3% | -0.34 | 0.73 | -1.49 | 2.22 | 18 |
| Melbourne | 219 | 226 | 49.2% | -0.33 | 0.81 | -1.31 | 2.12 | 17 |
| Collingwood | 247 | 257 | 49.0% | -0.45 | 0.99 | -1.01 | 2.00 | 11 |
| Richmond | 264 | 279 | 48.6% | -0.64 | 1.02 | -1.04 | 2.05 | 8 |
| Gold Coast | 271 | 291 | 48.2% | -0.84 | 1.19 | -1.04 | 2.23 | 12 |
| GWS | 282 | 305 | 48.0% | -0.95 | 0.82 | -1.22 | 2.04 | 16 |
| Geelong | 259 | 288 | 47.3% | -1.24 | 1.37 | -0.97 | 2.34 | 6 |
| Hawthorn | 300 | 334 | 47.3% | -1.35 | 1.39 | -1.11 | 2.51 | 1 |
| Brisbane | 250 | 283 | 46.9% | -1.43 | 0.81 | -1.39 | 2.20 | 15 |

Table 1: Centre Clearance Results by Team

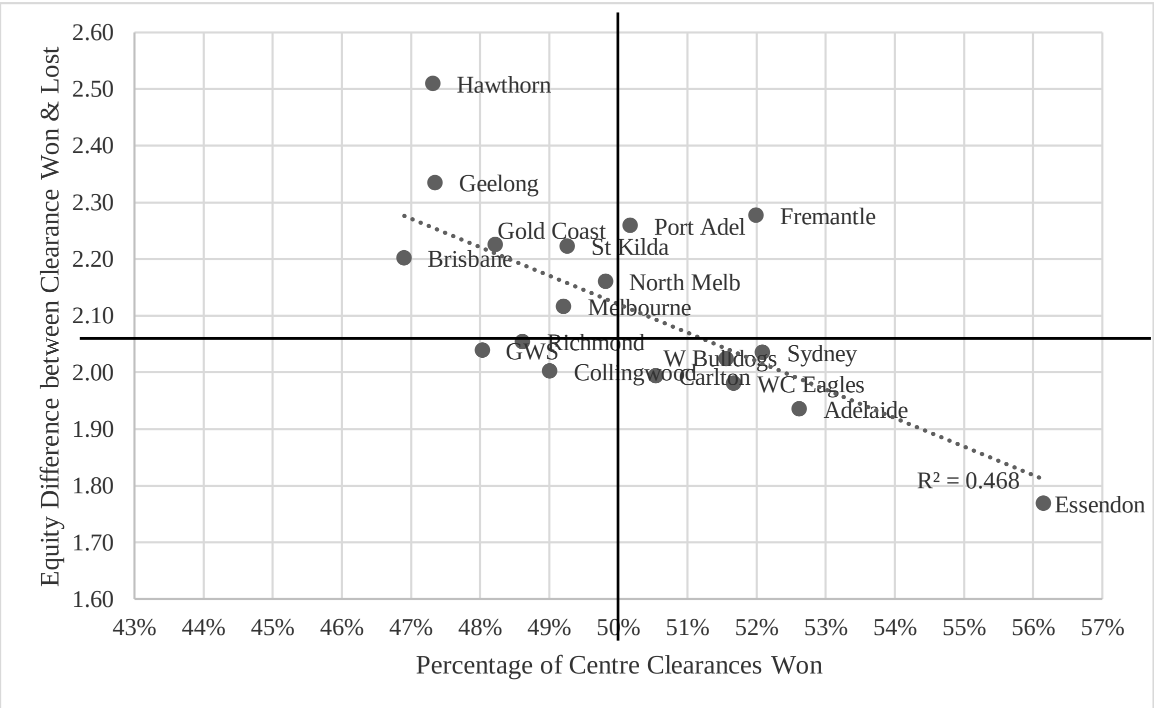



Figure 2: Equity Difference through Centre Clearance versus Centre Clearance Win%

There appears to be a strong negative correlation between winning a high percentage of Centre Clearances (Essendon being the outlier) and a reduced effectiveness of those opportunities as measured by equity. Additionally, there was no relationship between club quality as measured by their final position and any of the Centre Clearance indicators, apart from the obvious that equity metrics are higher for better teams.

MEASURING LUCK

A visualisation of the scoreboard margin as a sum of "mostly luck" and "mostly skill" effects has been developed. As an example, Table 2 shows the 2014 Preliminary Final which was won narrowly by Hawthorn, 15.7 (97) to 13.16 (94).

In this match, virtually all of the measurable luck-heavy indicators were in Hawthorn's favour, including conversion of scoring opportunities. The results of shots at goal have a heavy impact on the ultimate result, but are strongly dependent on random effects of the ball being dropped onto the swinging leg.

A traditional media analysis of this match would have emphasised the dominance at clearances and "clutch" shooting at goal as strengths of a team that went on to win the premiership the next week. Instead if we choose to monitor more stable indicators of on-field performance, our primary conclusion would be that Port Adelaide was very unlucky to lose.

| Mostly Luck | Raw | Equity |
|---|---|---|
| Centre Clearances | 16-12 | +4 |
| Ball-up Clearances | 11-8 | +3 |
| Throw-in Clearances | 21-14 | +7 |
| Shots at Goal | 15 / 23 | +17 |
| Oppo Shots at Goal | 13 / 31 | +5 |
| Unrealised Equity | | +2 |
| Mostly Skill | | |
| After Won Clearance | | -32 |
| After Lost Clearance | | -8 |
| After Kick-In | 7 / 15 | +1 |
| After Oppo Kick-In | 6 / 7 | +4 |

Table 2: Partition of Play by Phase

For a typical AFL match, using a continuous Markov process or approximating its absorbing states by a Poisson distribution, we can estimate the total variance of the final margin as approximately 1060 ($\sigma \cong 32.5$), varying slightly with the pace of the game. Approximately 34% of this variance is explained by the success or failure of shots on goal, and another 9% is explained by the clearance count. Considering the thousands of micro contests that happen around the ground and affect who emerges with the ball, this is a sizeable first step towards reducing the noise in the analysis of core performance.

## 4. DISCUSSION

The graph in Figure 2 caused surprise at Hawthorn FC, and requires substantial thought by experts in the sport as well as statisticians. The coaching group discussed what it meant if the centre clearance was effectively a coin flip, and how they might structure their defence, midfield and attack to respond to an event they have very little control over. The strong hint in the data that teams might over-invest in winning the clearance at the expense of resources in more impactful locations and roles is also a lesson to consider. Additionally, if the coaches cannot effectively intervene during the match to bring about a dominance of centre clearances – despite their obvious importance to the result – then they are free to stop worrying about that KPI during the game and focus their skills on pattern recognition and problem solving that computer algorithms cannot tackle.

Of course there is luck in every action on the field, and plenty of it in those phases labelled as "mostly skill". For any given contest, it is likely that luck dominates the difference in skill between two elite players. In any five-minute period, there is probably not enough data recorded to reliably inform decisions in the coaching box, yet almost every club speaks in terms of periods of dominance similar to this time-scale. Our obligation as statisticians is to patiently seek the signal in the data – quite the opposite of players who are having to continually react and physically respond in a complex environment. Coaches must straddle both camps, recognising their history as reactive players but forced to "think slow" and use their substantial knowledge of the game they are scrutinising to guide decisions. Classifying KPIs by importance to the result, and by the ability to reliably influence them, is a critical phase in sports analysis.

The sports analytics community has accepted measures like Expected Goals in recent years, recognising that the difference between a shot at goal succeeding or failing can be inches, and is well-described by stochastic models with sufficiently accurate data. These events happen at the end of a possession chain, and many people can now accept the epistemological argument that the difference between what *did* happen, and what *might have* happened (an array of counterfactuals) is a realisable metric.

It is harder to comprehend classifying events at the start of a possession chain in the same way, because we see a cascade of counterfactuals leading to an alternative reality. That way lies madness, we think, or at least chaos. Taking lessons from chaos theory, histories are expected to diverge at an exponential rate based on minuscule differences in initial conditions, and there are several decisions each second whence an alternative history could start its evolution.

The theory of backgammon offers a solution, as long as we are willing to briefly shut our eyes to the continuing play and just perform the mathematics, much like quantum physicists in the Feynman mould. The difference between the average path from what happened and the average path from what might have happened is expressible in scoreboard terms.

## 5. CONCLUSIONS

Lefgren, Platt & Price (2012) cleverly analysed American football data to discover that coaches would change their game plans in reaction to a loss, even if it was uninformative because they were expected to lose to that opponent. On the other hand a lucky or substandard win over an inferior team would not prompt as many changes. Sports analytics must become better at putting information in front of coaches after it has been normalised, had its variance reduced if possible, and is ordered by importance to success. Approaches such as that outlined in this paper could be applied to AFL and other sports, modifying the way data is collected, counterfactuals are considered, and reporting supports decisions.

### Acknowledgements

I wish to thank Hawthorn Football Club coaches and particularly David Rath, Head of Coaching Services, for their valuable discussions around this work.

### References

Gréhaigne, J.-F., & Godbout, P. (2011). Dynamic Systems Theory and Team Sport Coaching. Quest, 66 (1), 96-116.

Lefgren L, Platt B, Price J. (2012). Sticking with what (barely) worked. *NBER Working Papers № 17477*, National Bureau of Economic Research, Cambridge, MA.

Mauboussin, M.J. (2012). The Success Equation: Untangling Skill and Luck in Business, Sports, and Investing. Published by Harvard Business School Publishing.

O'Shaughnessy, D. (2012). The Luck of the Score: Partition of Variance for Performance Analysis of Team Sports. *Proceedings of the World Congress on Performance Analysis of Sport IX*, 37.

O'Shaughnessy, D. (2006). Possession versus Position: Strategic Evaluation in AFL. *Proceedings of the 8th Conference on Mathematics and Computers in Sport*. (Published in a condensed form in *Journal of Sports Science and Medicine*, 5, 533-540)

# CLUSTERING TEAM PROFILES IN THE AUSTRALIAN FOOTBALL LEAGUE USING PERFORMANCE INDICATORS

Bartholomew Spencer [a,b,d] , Stuart Morgan [c] , John Zeleznikow [a] , Sam Robertson [a,b]

[a] *Institute of Sport, Exercise and Active Living (ISEAL), Victoria University, Footscray, Australia*
[b] *Western Bulldogs Football Club, Footscray, Australia*
[c] *Australian Institute of Sport, Canberra, Australia*
[d] *Corresponding author:* Bartholomew.spencer@live.vu.edu.au

## Abstract

The relative importance of performance indicators to explain match outcome has been a recurring theme in team sports research. In Australian Rules football, prior studies have identified the relative importance of 'Inside-50's, kicks and goal conversion rates in explaining match success in the Australian Football League. However, it is possible there are different characteristics between teams that influence the relative importance of these variables. Therefore, this research aims to use unsupervised clustering methods to develop team profiles derived from summary statistics for each quarter of the game. Our data consists of commonly-reported performance indicators for all quarters played in the 2012, 2013 and 2014 AFL regular seasons. An initial random forest model was produced, from which measures of node purity revealed 'team' as an important variable in describing quarter outcome. A *k*-means approach was employed to classify quarters into *k*-types, independent of both team and outcome. This process revealed multiple winning methods across the range of score differences. Team profiles were then produced based on a team's *k*-type quarter frequencies, from which teams were clustered into distinct styles. Reintroducing outcomes revealed that similar profiles do not always produce similar outcomes, suggesting that execution is not always reflected in these discrete performance indicators. Finally, we explore if cluster centroids can be used to understand match-ups and we demonstrate how our classification approach can be applied in the longitudinal analysis of individual teams to identify the evolution of a team's strategy.

*Keywords: k*-means, Data mining, Australian football, Performance analysis, Opposition analysis

## 1. INTRODUCTION

The relative importance of performance indicators to match outcome has been a recurring theme in team sports research (Lupo, Condore & Tessitore, 2012; Sampaio, Lago, Casais & Leite, 2010), including Australian Rules football (Robertson, Back & Bartlett 2015; Stewart, Mitchell & Stavros, 2007). Prior studies have identified the relative importance of 'Inside-50's, kicks and goal conversion rates in explaining match success in the Australian Football League (AFL) (Robertson et al., 2015). This information is routinely used by clubs as a descriptive insight into those components of match play considered most crucial to success. Although some attempts have been made to look at network analyses, a lack of access to spatiotemporal data has limited investigations into how these performance indicators are determined. Consequently, information relating to the unique characteristics of different teams based on their player and ball movement remains limited. In using the information currently available, the dependence these discrete performance indicators have on team is yet to be investigated. Understanding whether success is achieved in the same way by all teams would allow for more informed coaching strategies to be developed. Based on heuristic match observations to date, it is hypothesised that team has a significant impact on the importance levels of different performance indicators. Hence, in this research we wish to demonstrate this importance before proposing a methodology for clustering what will be referred to as team profiles. In doing so, the effectiveness of profiles within the AFL will be analysed to identify if a singular profile is most explanatory of match outcome. Furthermore, applications of this work will be demonstrated via the longitudinal analysis of team profiles as a method for match preparation. Finally, contextual analysis of profile matchups will reveal if team profiles are consistent or if there exists fluidity amongst AFL teams.

## 2. METHODS

*DATA COLLECTION*

Data was compiled for all matches played during the 2012, 2013 and 2014 AFL regular seasons. This information was collected from Champion Data (CIA, Champion Data Pty Ltd, Southbank, Australia)[2] and included a total of 198 games for each 23-round season, producing a total sample size of 594 matches broken down by quarter. Each of the 2376 quarters had an entry for both teams, resulting in 4752 samples for the classification. For each match played during the three-year period, team totals for 22 commonly reported performance indicators were exported into Microsoft Excel. Data were recorded as the differences between the two teams, providing match context to the performance indicators. These were selected based on their inclusion in previous Australian Rules football (AF) studies which have demonstrated their relationship to outcomes (for examples, see Stewart et al., 2007; Robertson et al., 2015).

*RANDOM FOREST*

Data was split into training (80%) and testing (20%) sets, which were used to produce a model that explains *Margin* using the *Random Forest* (RF) algorithm. The RF algorithm works by producing *n* total decision trees, constructed from sampling the data, and produces an output equivalent to the mean of the total (Breiman, 2001). The purpose of this process was to inspect the ranking of *Team* amongst the variable importance plots, Figure 1, as an indication of its importance in explaining results. In particular, a plot of node purity, Figure 1b, indicated *Team* as the fifth most important variable in respect to reducing node impurities. In order to provide an example of the dependence of variable importance on *Team*, two individual RF models were produced, one for each of Richmond Tigers (RT) and Western Bulldogs (WB). A mean squared error (MSE) importance plot was produced for these models, from which variable rankings were used to identify any differences.

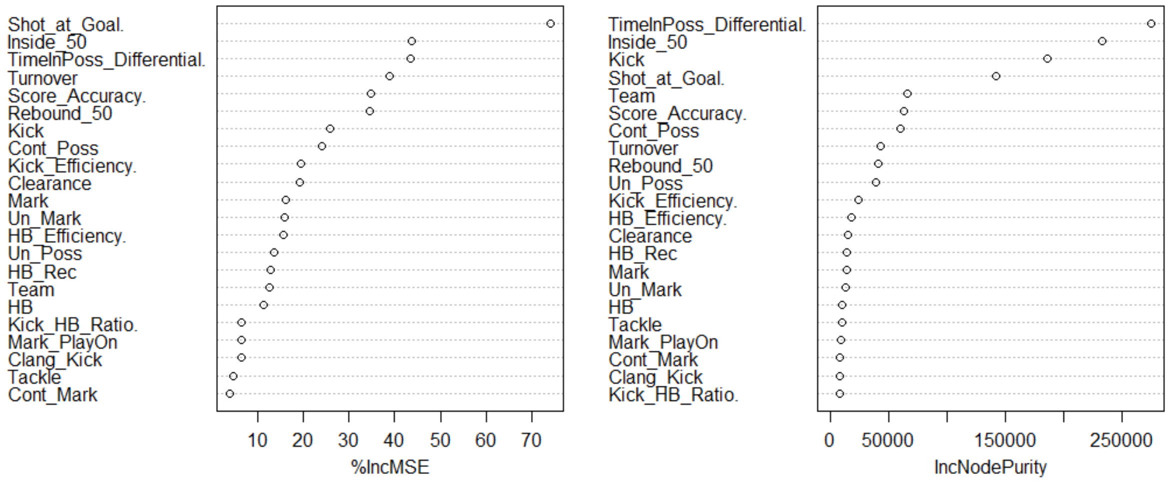



Figure 1. Variable importance plots for the RF model denoting (a) percentage increase in MSE if each variable were randomly permuted, and (b) the increase in node purities by splitting at each variable.

*QUARTER CLASSIFICATION*

A clustering approach was used to classify both team and quarter profiles across the dataset. *K*-means clustering, an unsupervised data mining technique, was chosen due to its demonstrated success in team sports literature (for examples, see Gyarmati et al., 2014). The *k*-means algorithm involves positioning *k* centroids repeatedly until equilibrium is reached when the within-cluster sum of squares (WCSS) reaches a minimum. This study employed a process of first classifying quarter types irrespective of *Team*. Thus, *Team* and *Margin* were removed from the analyses, before clustering the 4752 quarters into *k*-means clusters. The elbow method was used to select a *k*-value of 20. The elbow method involves visual inspection of the sum of squared error (SSE) as a function of *k*, identifying the preferred value for *k* as the value in which increasing *k* yields less improvement to SSE – i.e. the point of diminishing returns (Kodinariya & Makwana, 2013). To identify the presence of the most successful winning strategies in the AFL, *Margin* was reintroduced to the dataset and the average *Margin* (mean ± standard deviation) was calculated for each cluster.

[2] https://www.championdata.com/ last accessed May 11 2016.

*TEAM CLASSIFICATION*
Team profiles were defined as a team's frequency of each $k$-means quarter. Thus, a team's profile is comprised of 20 variables, each corresponding to the frequency by which they experienced that quarter type. To allow for classification regardless of sample size, these variables were normalised as a percentage of their total quarters. Teams were classified into eight clusters, chosen via the elbow method, which we will refer to as the eight team profiles present in the AFL. As such, two teams with similar quarter clusters can be part of the same team profile, suggesting similarities in how they have achieved success in regards to their performance indicator compositions. As was the case with the quarter clusters, average margin (mean ± standard deviation) was calculated for each of the team clusters. While teams could have been clustered via mean performance indicators across the three seasons, choosing to cluster via $k$-means quarter cluster frequencies ensured variation was accounted for and required no assumptions about variable distributions.

To demonstrate the tactical applications of this study, the $k$-means centroid equation was derived from our team classification. Classifying a team into one of the existing team profiles involves calculating the Euclidean distance between a team's quarter profiles, $X$, and each of cluster centroids, $\mu_s$, the minimum of which determines a team's classification (see Equation 1). We demonstrate the longitudinal analysis of a team through manually classifying individual seasons of Port Adelaide (PA), a team that experienced a change in coach during the seasons in this study (prior to the commencement of the 2013 season). PA were classified for each individual season to identify any changes in team profile in response to coaching methods. In doing so, this study aimed to demonstrate the application of this methodology for opposition analysis.

$$\underset{\{s|s\in\mathbb{N},s\geq k\}}{\arg\min}\ f(s)\|X-\mu_s\|^2 \tag{1}$$

To explore the presence of transitive relationships in the AFL (i.e., whether a team's profile is consistent or if it has a dependent relationship on their opponent's profile), matchup profiles were created as the frequency of quarter types experienced by a team of *Profile A* when competing against a team of *Profile B*. From this, we classify a new profile, *Profile AB*, using the centroid equation, allowing for insights into cluster matchups that may inform match preparation.

## 3. RESULTS

*RANDOM FOREST*
Results of the individual RF models, in the form of an increase in MSE plot, are presented in Figure 2. Notably, *Inside_50*s had greater importance for WB compared to RT, while RT had higher values than WB for the possession and accuracy score accuracy variables. Furthermore, while WB had higher values for *Kick_HB_Ratio* and *HB_Efficiency*, the remaining handball-related variables (*HB* and *HB_Rec*) had greater importance in the RT model. Mean squared errors for the individual models revealed that the RT model (*RMSE* = 8.06) was able to explain margin more effectively than the WB model (*RMSE* = 9.85).

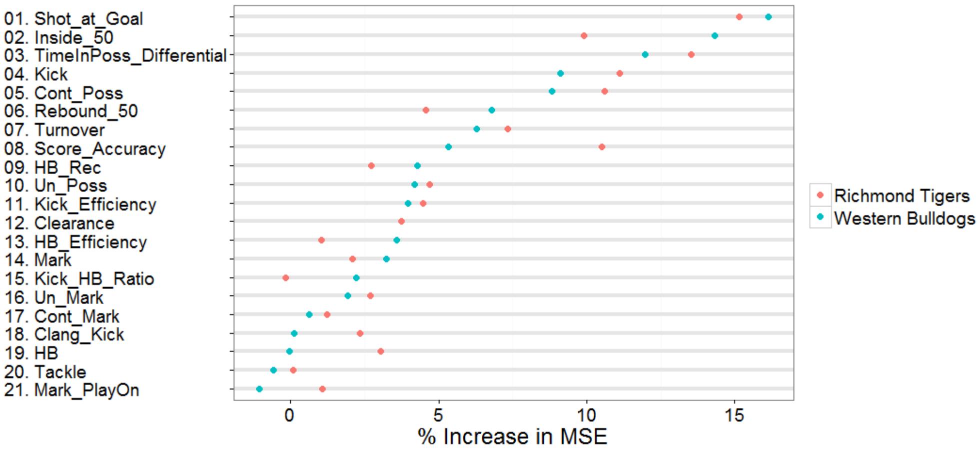



Figure 2. Percentage increase in MSE plot for the individual RF models. Note that performance indicators are ranked according to their relative importance for WB.

*QUARTER CLASSIFICATION*

Summary statistics for the quarter clusters are presented in Table 1. The table reveals multiple quarter types at all ranges of *Margin*. Furthermore, there exists no cluster that achieved complete success or failure, with all containing winning and losing quarters. From these results, it can be noted that Clusters 4, 10, 11, 15 and 20 produced the most favourable average *Margin*, while 3, 8, 9 and 14 produced the least favourable average *Margin*.

Table 1. The *Margin* means and standard deviations for Quarter Clusters

| Cluster ID | *Margin* mean | *Margin* s.d. | % *Wins* | Cluster ID | *Margin* mean | *Margin* s.d. | % *Wins* |
|---|---|---|---|---|---|---|---|
| 1 | 3.44 | 13.11 | 60.0% | 11 | 17.24 | 12.47 | 88.3% |
| 2 | -5.65 | 12.33 | 30.3% | 12 | 9.41 | 11.69 | 75.7% |
| 3 | -25.64 | 14.81 | 1.3% | 13 | -3.77 | 12.58 | 35.4% |
| 4 | 18.54 | 12.98 | 92.9% | 14 | -21.09 | 12.34 | 4.6% |
| 5 | -1.04 | 13.35 | 46.6% | 15 | 13.21 | 11.9 | 87.6% |
| 6 | 2.16 | 12.41 | 53.6% | 16 | -2.08 | 12.95 | 40.0% |
| 7 | 0.24 | 13.11 | 47.1% | 17 | -7.43 | 12.07 | 23.0% |
| 8 | -19.16 | 14.25 | 8.2% | 18 | -7.77 | 12.52 | 28.5% |
| 9 | -18.63 | 11.57 | 5.7% | 19 | 4.82 | 13.21 | 66.5% |
| 10 | 26.15 | 15.64 | 95.9% | 20 | 20.48 | 12.73 | 92.8% |

*TEAM CLASSIFICATION*

A plot of the teams in respect to their respective mean and standard deviation of *Margin* is presented in Figure 3, with summary statistics recorded in Table 2. Notably, the only team that occupied its own profile (i.e., a cluster containing only one object) was Geelong. Geelong's close proximity (in regards to *Margin*) to multiple teams of different profiles enforces the notion that profile is dependent on more than outcome. Inspection of teams located below the *x*-axis origin (i.e. a mean *Margin* below 0) reveals that clusters 5 and 6 are dominated by the other four, with no team of either cluster producing favourable *Margin* statistics. These clusters are comprised of a high number of Type 2, 13 and 14 quarters, which correspond to high handball variables, low contested variables, and low shot accuracy variables respectively.

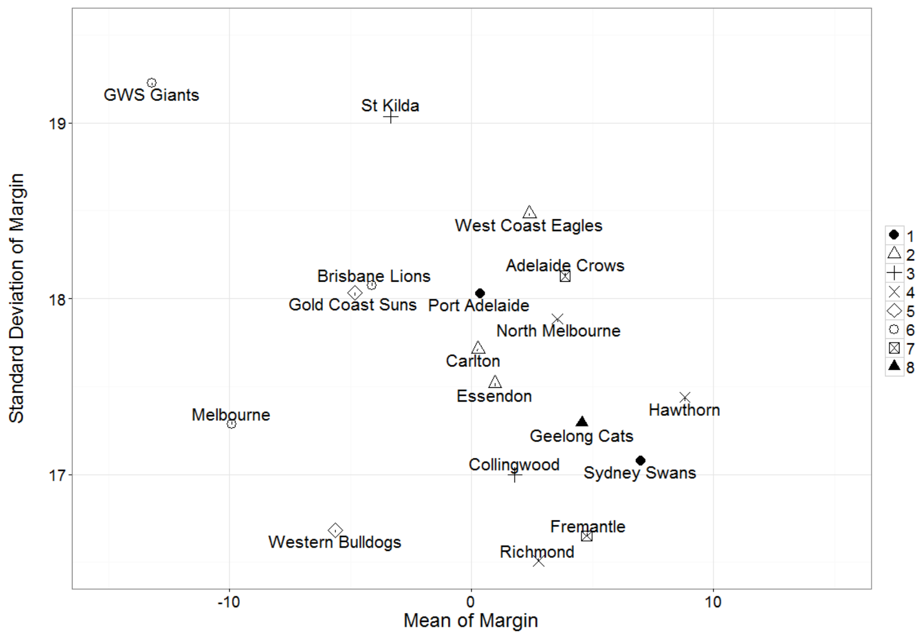


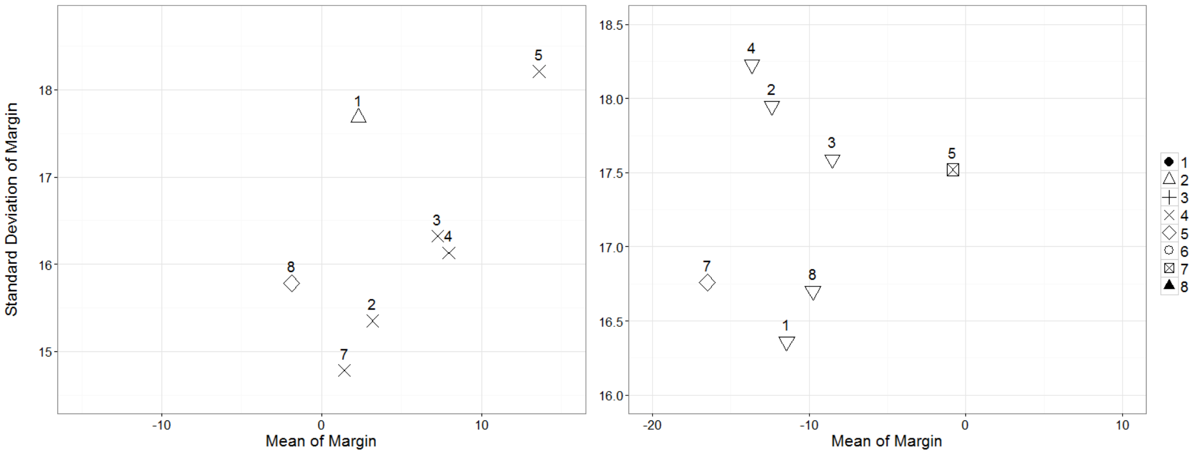

Figure 3. Scatterplot of the 18 AFL teams with respect to their *Margin* mean and standard deviation, with shapes corresponding to their *k*-means cluster.

Individual classification of the three PA seasons revealed a change in classification from *Profile 1* in 2012 and 2013, to *Profile 3* in 2014. Analysis of these individual seasons revealed the 2014 season to have the highest mean *Margin* and lowest standard deviation of the three seasons. These findings are noteworthy given the average profile statistics, Table 2, which suggests teams categorised in *Profile 1* generally outperform those categorised in *Profile 3*. To test if this variation was inherent in all teams, this classification process was tested on WB, a team who experienced no change in coach across the three seasons, who was found to be consistently classified as *Profile 5*.

Table 2. The *Margin* means and standard deviations for the eight Team Profiles

| Profile ID | *Margin* mean | *Margin* s.d. | % *Wins* |
|---|---|---|---|
| 1 | 3.69 | 17.85 | 57.8% |
| 2 | 1.21 | 17.91 | 49.7% |
| 3 | -0.78 | 18.21 | 45.1% |
| 4 | 5.03 | 17.47 | 60.7% |
| 5 | -5.23 | 17.36 | 36.7% |
| 6 | -9.08 | 18.57 | 29.7% |
| 7 | 4.30 | 17.39 | 58.5% |
| 8 | 4.55 | 17.29 | 60.2% |

Results of the style matchups are presented in Figure 4. These figures denote the contextual profiles of teams in *Profile 4* (Figure 4a) and *Profile 6* (Figure 4b). Based on these results it can be noted that the two matchups that *Profile 4* produced different performance indicator profiles against were amongst those that they performed poorly against. For *Profile 6*, reclassification occurred in their best matchup (*Profile 5*) and their worst matchup (*Profile 7*).




(a) *Profile 4* (b) *Profile 6*

Figure 4. Plot of individual matchups for (a) *Profile 4*, and (b) *Profile 6*, with labels corresponding to the opposing *Profile.*

## 4. DISCUSSION

This study has demonstrated the importance of *Team* in respect to quantifying the relative importance of performance indicators in the AFL. Furthermore, we have demonstrated a method for identifying similarities between these teams via the development of team profiles using a repeated clustering approach. This methodology clusters teams regardless of the distribution and averages of their performance indicator differences, thus accounting for the large variations when using relative values.

The results of the individual RF models revealed that a significant difference is observed in both the variable importance and the model accuracy (*RMSE*). This second point in particular suggests that, between teams, performance indicators capture different amounts of success, thus, depending on the team that is analysed, *Margin* can be predicted with varying accuracy. This reveals that some teams are more consistent in their use of different variables, in terms of how these variables relate to outcomes – i.e., that one team may execute kicking strategies more consistently, hence kicking related variables would have more impact on the RMSE of predictive models.

Classifying quarters into 20 *k*-means clusters revealed that, regardless of the makeup of a cluster, execution is not always captured via performance indicators in AF. Furthermore, while some clusters are objectively preferable, the existence of multiple winning types should be noted. Breaking down the dominant clusters (Clusters 4, 10, 11, 15 & 20) reveals distinct differences in their performance indicator makeup. For example, Cluster 4 is comprised of high handball related variables and low kicking, while Cluster 20 consists of high kicks, marks and goal accuracy. The similar success of these two clusters suggests that a single optimal strategy does not exist in the current state of AF.

As was the case with quarters, the *Team* classification revealed the presence of multiple successful profiles. Notably, the top three teams (in regards to *Margin* statistics) were classified into different profiles, and each profile was comprised of teams of varying success in most cases. This enforces the notion that execution is not completely captured via performance indicators in their current state. The practical applications of this study were demonstrated through longitudinal and contextual (Figure 4) analysis of these team profiles. Contextual analysis revealed that profile matchups result in a variety of profile classification, suggesting that, regardless of ladder positions, gameplay may be largely responsive. While this may not be a new notion, our methodology provides statistical examples of this.

One notable limitation of this study is the use of performance indicator differences for the analysis. The use of differences is a step towards providing match context above the use of raw values however it may not capture important attributes such as the percentage differences, the pace of the match or consideration of sequences of events in play. Furthermore, given the nature of these statistics we can not necessarily correlate these profiles to game styles in their current state. However, the use of methodology that isn't specific to data allows for these profiles to be developed regardless of our data source. For example, GPS data could be substituted for similar classification of quarter types, hence team profiles, to cluster teams based on movements were rich enough data available.

## 5. CONCLUSIONS

Prior studies into the relative importance of performance indicators have focused analyses on the playing group as a whole, rather than researching this importance as a function of *Team*. Our research has demonstrated the contextual importance of these performance indicators in the AFL via the development of predictive modelling. Team profiles were produced via consecutive *k*-means clustering and revealed that, in the current state of the AFL, success is achieved by different methods. Furthermore, longitudinal analysis of team classification showed that this methodology is capable of capturing a change in coaching style (for example Port Adelaide) and contextual analysis of these matchups suggested that a team's profile is dependent on their opponent.

**Acknowledgements**

We wish to thank Champion Data for providing access to the data used in this study.

**References**

Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.

Gyarmati, L., Kwak, H., & Rodriguez, P. (2014). Searching for a Unique Style in Soccer. arXiv preprint arXiv:1409.0308.

Lupo, C., Condello, G., & Tessitore, A. (2012). Notational analysis of elite men's water polo related to specific margins of victory. *Journal of Sports Science & Medicine*, 11, 516.

Kodinariya, T., & Makwana, P. (2013). Review on determining number of cluster in *K*-means clustering. *International Journal of Advance Research in Computer Science and Management Studies*, 1(6), 90-95.

Robertson, S., Back, N., & Bartlett, J. D. (2015). Explaining match outcome in elite Australian Rules football using team performance indicators. *Journal of Sports Sciences*, 1-8.

Sampaio, J., Lago, C., Casais, L., & Leite, N. (2010). Effects of starting score-line, game location, and quality of opposition in basketball quarter score. *European Journal of Sport Science*, 10, 391-396.

Stewart, M., Mitchell, H., & Stavros, C. (2007). Moneyball applied: Econometrics and the identification and recruitment of elite Australian footballers. *International Journal of Sport Finance*, 2(4), 231-248.

# WHY DO COLLINGWOOD HAVE SO MANY FATHER-SON SELECTIONS? A DEMOGRAPHIC ANALYSIS OF THE AFL'S FATHER-SON RULE

Geoffrey N. Tuck [a,b]

[a] *CSIRO Data 61 and CSIRO Oceans and Atmosphere*
[b] *Corresponding author: geoff.tuck@csiro.au*

## Abstract

The Australian Football League's (AFL's) Father–Son rule is a unique player drafting rule that allows sons of former players to be selected by their father's club. The rules that determine eligibility have undergone numerous changes since its introduction in 1949, including rules for new teams from outside of the traditional Victorian-based (VFL) clubs that had no history of fathers from which Father–Son selections could be derived. The observed number of Father–Son selections to each club is markedly different between the Victorian-based clubs, and between the Victorian and non-Victorian-based clubs. In this paper, a demographic model and player data from the AFL and the state leagues (WAFL and SANFL) are used to estimate the annual number of available sons to each of the AFL clubs. Results show that the observed number of selections can largely be explained by the number of available sons. The relatively large number of Father-Son selections to the Collingwood and Geelong Football Clubs (12 and 11 selections since 1986 respectively) coincides with them having a larger number of available sons than say the St. Kilda Football Club (3 selections). In addition, the model was able to provide an explanation for the lower number of Father-Son selections to the non-VFL aligned clubs in comparison to the VFL-aligned clubs by linking the lower estimated number of expected sons to the rules that were in place for these sides at the time. In particular, the Adelaide Crows (0 selections) and Fremantle Dockers (1 selection) have had nearly half the expected number of available sons compared to the Victorian-based clubs since their establishment. The model can also be used to predict the number of available sons into the future, and so can be used to guide management decisions regarding competitive balance if further modifications to the AFL's Father–Son rule are required.

**Keywords**: Player draft, demographic modelling, competitive balance

## 1. INTRODUCTION

The Australian Football League's (AFL's) Father-Son (FS) rule provides an opportunity for the sons of former players to preferentially be selected by the father's club (Borland, 2006; Tuck, 2015). Introduced to the Victorian Football League (VFL) in 1949, the FS rule is designed to promote family traditions and allows a small degree of club choice beyond the more constraining reverse-order AFL player draft that was introduced in 1986 (Tuck et al., 2015). However, the rule also has led to contention due to the apparent inequity in clubs' ability to select players through the FS rule (Rucci, 2009; Niall, 2012; Quayle, 2014). Consideration of the number of FS selections by club shows a broad distribution; from Collingwood and Geelong having 12 and 11 selections respectively, to Adelaide, Fremantle and Port Adelaide having one or no selections (Table 1). If an implicit objective of the FS rule is to maintain equitability across clubs (through an equal probability of selection), then Tuck (2015) showed that a test of the assumption of an equal proportion of successes (Father–Son selections) across all teams (accounting for years in the competition) leads to rejection of this hypothesis.

The general principle underlying the FS rule is that if a player has played sufficient senior games for his club, then the player's son has the opportunity to be selected by the father's club. The number of qualification games and the sacrifices (through the draft) in order to obtain the FS selection have changed markedly over time (AFL, 2015; Tuck, 2015). In addition, rules were devised to allow clubs outside of Victoria, with no history of players, an opportunity to draft sons of fathers from their local state leagues. The non-VFL aligned clubs are West Coast (established 1987), Fremantle (1995) with local league the Western Australian Football League (WAFL), and Adelaide (1991), Port Adelaide (1997) with local league the South Australian National Football League (SANFL). The Brisbane Lions (via Fitzroy) and Sydney Swans (via South Melbourne) operate under the same qualification rules as the VFL-aligned clubs. The different rules for VFL-aligned and non-VFL aligned clubs immediately introduces the prospect that these rules could lead to differences in a club's ability to select players under the FS rule. Examination of the selections by club (Table 1) shows that three of the four non-VFL aligned clubs have had the lowest number of FS selections. The current AFL rule for FS eligibility requires that the father has played 100 or more senior games for his VFL/AFL club (Table 2 and Table 1a of Tuck, 2015). The minimum number of games for qualification has been as low as 20 (Table 3). The initial FS rule for the non-VFL aligned clubs followed the minimum games requirement of the VFL aligned clubs, with the additional need for the player to have played at least one game in their local state-based league. In 2001 this rule was removed and replaced by a state league based qualification. For the WA-based and SA-based clubs, fathers must have played 150 senior WAFL games or 200 senior SANFL games prior to the establishment of the AFL club. The 100 game AFL rule also applies (Table 3).

Tuck (2015) showed that a more appropriate way of exploring the disparity of Father–Son selections than considering selections by year is to consider the expected number of sons that are available to each of the clubs. The number of fathers that qualify will differ between clubs and, depending on their age, the expected number of sons to those fathers will differ over time. Using a demographic model and updated player data from the VFL/AFL, this paper refines and updates the model of Tuck (2015). Further details of the model, data used, qualification rules and history of the Father-Son rule can be found in Tuck (2015). This paper also corrects Rule 12 of Table 1a of Tuck (2015) where the years over which the games qualification for WA and SA clubs differed (see Table 2). This change made only minor differences to the resulting estimated expected sons, and does not change the conclusions of that paper.

| Club | Number of selections | Seasons since 1986 or from establishment | Number per season |
|---|---|---|---|
| Collingwood | 12 | 30 | 0.400 |
| Geelong | 11 | 30 | 0.367 |
| Carlton | 9 | 30 | 0.300 |
| Essendon* | 7 | 30 | 0.233 |
| Western Bulldogs | 7 | 30 | 0.233 |
| Melbourne | 6 | 30 | 0.200 |
| Richmond | 6 | 30 | 0.200 |
| **West Coast** | **5** | **29** | **0.172** |
| Sydney Swans | 5 | 30 | 0.167 |
| Brisbane | 3 | 29 | 0.103 |
| Hawthorn | 3 | 30 | 0.100 |
| St. Kilda | 3 | 30 | 0.100 |
| North Melbourne | 2 | 30 | 0.067 |
| Fitzroy | 1 | 9 | 0.111 |
| **Port Adelaide** | **1** | **19** | **0.053** |
| **Fremantle** | **1** | **21** | **0.048** |
| **Adelaide** | **0** | **25** | **0** |

Table 1: Father-son selections by club. The number of Father-Son selections by each AFL club between 1986 and 2015, the number of seasons each club could have had a Father-Son selection since 1986, and the number of selections per season. The non-VFL aligned clubs are in bold. Brisbane refers to a combination of the Brisbane Lions and Bears. * includes two rookie Father-Son selections.

## 2. METHODS

### PLAYER DATA

The number of games played is the key data requirement for an exploration of the number of players that qualify under the FS rule. Player data (games, birth year, state of origin) were obtained from the AFL for all players between 1897 and 2015 (source: AFL, March 2016). Data for the WAFL and the SANFL were obtained from those leagues and various club historians and websites.

### PATERNITY RATES

To estimate the expected number of sons by eligible player and year, data on paternity rates are needed. Age-specific paternity rates (ASPR) are not publicly available. As such, ASPR was estimated from annual age-specific fertility rate (ASFR) data from 1921 to 2014 obtained from the Australian Bureau of Statistics (ABS). ASFR was transformed to ASPR using the ratio of the known paternity to fertility rates from 2014 (ABS, 2015a,b; Tuck, 2015).

| FS Rule # | First year of application | Rule and amendments |
|---|---|---|
| 11 | 2001 | The clause that allowed sons of the President, Vice President, General Manager, Senior Coach or Elected Member of the Senior Grade Committee of any Club was deleted. |
| 12 | 2001 | A State League games qualification was introduced for WA (150 WAFL senior games) and SA (200 SANFL senior games) Clubs. These games must have been played during the period 1967-1987 for West Coast, 1975-1995 for Fremantle, 1970-1990 for Adelaide (note Adelaide began in 1991) and 1977-1997 for Port Adelaide. Included a "sunset" clause noting this provision would be in place for a 20 year period from the time the particular Club entered the competition. |
| 13 | 2003 | The VFL/AFL qualification increased to 100 games. |
| 14 | 2003 | The QAFL and SFL eligibility clauses were removed due to the Sydney/South Melbourne, and Brisbane Lions/Brisbane Bears/Fitzroy father-son provisions being more relevant and applicable |
| 15 | 2006 | The WA and SA AFL Clubs 150 and 200 State League games qualification was amended with the "sunset" clause replaced with a provision noting the father of the player must have played the relevant home and away and finals series matches prior to but not including the year the Club entered the AFL competition. |

Table 2: The Father-Son qualification rules since 2001 with their year of first application (R. Austin, pers. comm.; Guide to the AFL Player Draft System, and AFL National Draft pamphlets). Rule 12 has been amended from Tuck (2015). For the full set of qualification and selection rules, see Table 1 of Tuck (2015).

## MODEL DESCRIPTION

A detailed description of the model equations can be found in Tuck (2015) and is not repeated here. However, essentially the model follows the steps below:

1. Determine the eligibility of every player in the VFL/AFL, SANFL and WAFL, given a Father-Son rule, *s*.
2. The birth year of an eligible father *f* is then used in combination with the ASPR and annual sex-ratios (ABS, 2015a) to determine the expected number of sons in a particular year *y* if the father is of age *a*.
3. The expected number of sons is then translated forward 18 years (the drafting age), assuming negligible mortality between birth and 18 years of age.
4. The number of expected sons in a particular year for club *c* under Father–Son rule *s* is then the sum over each of these expectations for each eligible father from each club over all child-bearing ages of the father (ages 15–50), where account is taken of multiple club eligibility (the expectation is divided between clubs).
5. As there have been several changes to the Father-Son rules, the expectations are calculated for each rule and then combined in accordance with the years over which the rule applies.

The procedure is illustrated in Figure 1.

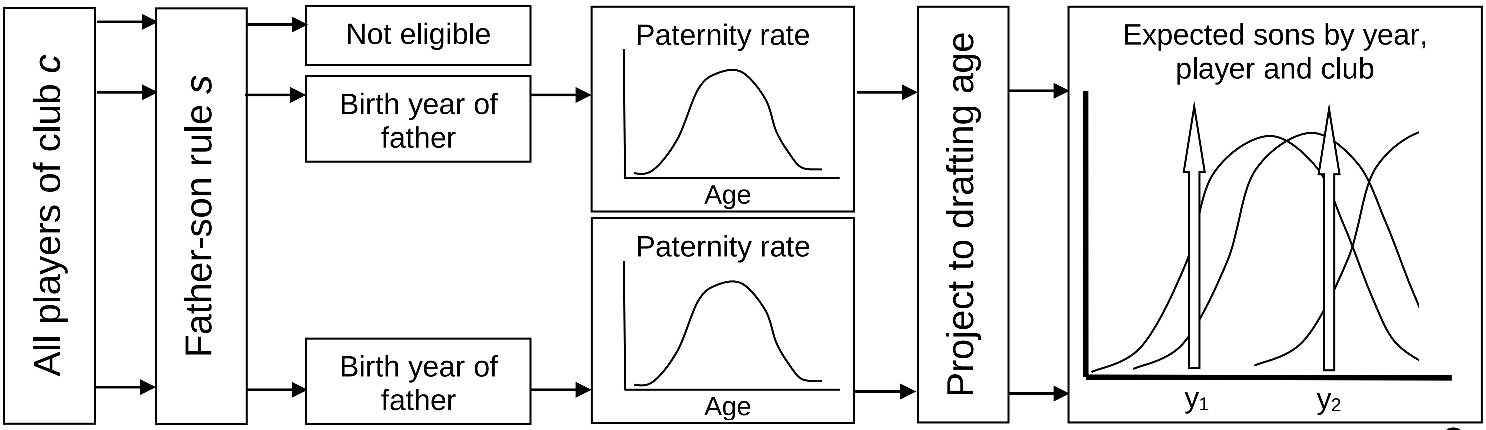



Figure 1: A diagrammatic representation of the demographic model. The hollow arrows in the right hand box represent the sum of expected sons across all eligible fathers of a club for a particular year.

## 3. RESULTS

- The annual number of expected sons by club changes markedly as eligibility rules change (Table 3).
- Large differences in annual expected sons exist between VFL-aligned clubs (Richmond, St Kilda, Sydney less than Hawthorn and Essendon).
- Differences also exist between non-VFL aligned clubs, with Adelaide having less expected sons than Port Adelaide (likely due to Adelaide having 2 less associated SANFL clubs; Tuck, 2015), and Fremantle less than West Coast.
- Longitudinal comparisons (Table 4) between VFL-aligned clubs show that since 1986, Collingwood, Hawthorn and Geelong have had the greatest number of expected sons, while St. Kilda and Melbourne have had the least (see the ‘Since 1986 (VFL)’ column).
- For the non-VFL aligned clubs, the expected number of sons following their establishment is considerably lower than the VFL-aligned clubs (eg see the ‘Since 1991 (ADEL)’ column of Table 4).
- Following the introduction of the state-based FS rules (in 2001), this disparity has largely disappeared (see the ‘Since 2001’ column of Table 4).
- Projections of expected sons from 2016 to 2025 show that West Coast, Melbourne and Essendon have the largest values, while St. Kilda and Richmond have the lowest expected sons over this period (Table 3).

## 4. DISCUSSION

The AFL’s Father-Son rule is a unique (among major sporting leagues) player drafting rule that allows the romanticism of family traditions to continue within a club (Tuck, 2015). However, the rule has also been controversial as it can override (AFL, 2015) the basic reverse-order annual player draft that may have allowed the player to be drafted by an alternate, possibly lower-ranked, club. In addition, there is also an apparent disparity in the ability of clubs to obtain players through this rule (Table 1; Niall, 2012; Quayle, 2014; Rucci, 2009). Since 1986, Collingwood and Geelong have had the most selections (with 12 and 11 selections each), while North Melbourne, Port Adelaide, Fremantle and Adelaide have had the least (with 2, 1, 1 and 0 respectively). As a consequence, the recent successes of the Collingwood and Geelong Football Clubs have been linked to their high quality Father-Son selections (Niall, 2012; Anderson, 2013).

To explore the apparent broad distributions of selections, Tuck (2015) introduced a demographic model that determined player eligibility for every VFL/AFL, SANFL and WAFL player under the various FS rules, and then projected paternity rates to produce an expected number of sons for each player, each club and each year. This paper has updated that work. Results here (and in Tuck (2015)) showed that the expected number of sons by club generally matched the actual observations of selections. This indicates that the number of selections by club are, not surprisingly, a function of the number of eligible fathers (i.e. the number of players having played more than the minimum required games and the FS rule that applied at the time). While outside the scope of this analysis, it should also be recognised that there may also be preferential selection of clubs by eligible players that strongly desire to be at certain clubs (having been fostered through club Father-Son academies for example).

The lower number of expected sons to the non-VFL aligned clubs following their establishment is likely due to the initial FS rules where eligible players for a non-VFL aligned club were also eligible for VFL-aligned clubs (thus halving, at least, the expected sons). If a more equitable distribution of expected sons (since establishment of the non-VFL aligned clubs) is a desirable management objective, then this model could be used to explore alternative minimum game thresholds. Namely, the 150 and 200 game threshold for the WAFL and SANFL could be reduced, or the 100 game threshold for the VFL-aligned clubs could be increased. However, an analysis of the labour regulations for the non-VFL aligned clubs should also recognise that these clubs were provided additional assistance during their establishment, e.g. in the form of moratoriums on drafting players from WA and SA for the clubs outside of these states (Macdonald and Booth, 2007). In addition, it is also worth noting that since the state-based rules were introduced in 2001 (Table 2, Rule 12), the expected number of sons for the non-VFL aligned clubs has increased (Table 4; ‘Since 2001’ column). As stated by Tuck (2015), this does not appear to have increased the number of selections for the non-VFL aligned clubs, with only three state-based selections (players Peake, Ebert, Morton) to the four non-VFL aligned clubs since 2001, in comparison to 39 selections for the 12 VFL-aligned clubs over the same period.

| Minimum VFL/AFL games | 20 | | | | | | | 50 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| WA/SA Clubs | | Club qualification through minimum games rule and at least one game in the local league | | | | | | | | | | | | | |
| | 1986 | 1987 | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 |
| ADEL | | | | | | 0.6 | 0.6 | 0.3 | 0.4 | 0.4 | 0.4 | 0.3 | 0.3 | 0.4 | 0.4 |
| BRIS | | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 2.3 | 2.3 | 2.4 | 2.3 | 2.3 |
| CARL | 4.8 | 4.8 | 4.8 | 4.8 | 4.3 | 4.1 | 3.7 | 2.2 | 2.0 | 2.0 | 2.0 | 2.0 | 1.9 | 1.8 | 1.9 |
| COLL | 5.3 | 5.3 | 5.1 | 5.2 | 4.9 | 4.4 | 4.2 | 2.5 | 2.5 | 2.3 | 2.2 | 2.2 | 2.1 | 2.2 | 2.2 |
| ESS | 4.4 | 4.6 | 4.4 | 4.8 | 4.5 | 4.0 | 3.8 | 2.2 | 2.3 | 2.3 | 2.2 | 2.1 | 2.2 | 2.3 | 2.4 |
| FITZ | 5.0 | 4.9 | 4.8 | 5.0 | 4.7 | 4.3 | 4.0 | 2.5 | 2.3 | 2.2 | | | | | |
| FREM | | | | | | | | | | 0.7 | 0.7 | 0.7 | 0.7 | 0.8 | 0.8 |
| GEEL | 5.6 | 5.5 | 5.4 | 5.5 | 5.0 | 4.4 | 4.1 | 2.3 | 2.2 | 2.2 | 2.1 | 2.1 | 2.0 | 2.1 | 2.1 |
| HAW | 5.1 | 5.0 | 4.9 | 5.0 | 4.7 | 4.1 | 3.7 | 2.4 | 2.2 | 2.2 | 2.2 | 2.2 | 2.1 | 2.1 | 2.1 |
| MELB | 4.7 | 4.7 | 4.7 | 4.8 | 4.6 | 4.2 | 3.8 | 2.0 | 1.9 | 1.8 | 1.8 | 1.7 | 1.7 | 1.8 | 1.9 |
| NTH | 4.9 | 4.7 | 4.8 | 5.2 | 4.9 | 4.1 | 3.7 | 2.5 | 2.6 | 2.5 | 2.4 | 2.2 | 2.2 | 2.2 | 2.1 |
| PORT | | | | | | | | | | | | 0.3 | 0.3 | 0.3 | 0.4 |
| RICH | 5.4 | 5.3 | 5.2 | 5.3 | 4.9 | 4.4 | 4.0 | 2.2 | 2.1 | 2.0 | 2.0 | 1.9 | 1.9 | 1.9 | 1.9 |
| STK | 4.8 | 4.7 | 4.8 | 4.8 | 4.3 | 3.9 | 3.6 | 2.4 | 2.2 | 2.1 | 2.1 | 2.1 | 2.0 | 2.0 | 2.0 |
| SYD | 5.7 | 5.6 | 5.7 | 5.8 | 5.4 | 4.5 | 4.1 | 2.3 | 2.3 | 2.3 | 2.2 | 2.1 | 2.1 | 2.1 | 2.2 |
| WESTC | 0.0 | 1.1 | 1.1 | 1.2 | 1.2 | 1.2 | 1.2 | 1.0 | 1.0 | 0.7 | 0.7 | 0.8 | 0.8 | 1.0 | 1.0 |
| WsB | 5.5 | 5.3 | 5.2 | 5.3 | 5.1 | 4.6 | 4.1 | 2.2 | 2.1 | 2.0 | 1.9 | 1.9 | 2.0 | 2.1 | 2.0 |

| Minimum VFL/AFL games | 50 | | 100 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| WA/SA Clubs | State-league eligibility through the WAFL/SANFL | | | | | State-league eligibility amended | | | | | | | | | | |
| | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 2025 |
| ADEL | 1.3 | 1.3 | 1.4 | 1.4 | 1.3 | 1.4 | 1.4 | 1.4 | 1.4 | 1.3 | 1.2 | 1.3 | 1.3 | 1.3 | 1.2 | 13.5 |
| BRIS | 2.5 | 2.5 | 1.2 | 1.1 | 1.2 | 1.2 | 1.2 | 1.3 | 1.4 | 1.4 | 1.4 | 1.4 | 1.4 | 1.5 | 1.5 | 15.7 |
| CARL | 2.2 | 2.1 | 1.4 | 1.3 | 1.4 | 1.5 | 1.4 | 1.5 | 1.4 | 1.5 | 1.5 | 1.4 | 1.4 | 1.4 | 1.4 | 14.4 |
| COLL | 2.3 | 2.1 | 1.4 | 1.4 | 1.5 | 1.4 | 1.4 | 1.5 | 1.4 | 1.5 | 1.5 | 1.5 | 1.5 | 1.4 | 1.4 | 14.4 |
| ESS | 2.5 | 2.3 | 1.3 | 1.3 | 1.3 | 1.3 | 1.4 | 1.5 | 1.4 | 1.5 | 1.4 | 1.5 | 1.5 | 1.5 | 1.5 | 16.2 |
| FITZ | | | | | | | | | | | | | | | | |
| FREM | 0.7 | 0.7 | 0.9 | 0.9 | 0.8 | 1.4 | 1.4 | 1.5 | 1.4 | 1.4 | 1.3 | 1.3 | 1.4 | 1.4 | 1.3 | 13.7 |
| GEEL | 2.2 | 2.2 | 1.4 | 1.3 | 1.3 | 1.2 | 1.3 | 1.4 | 1.3 | 1.4 | 1.4 | 1.4 | 1.4 | 1.3 | 1.4 | 15.0 |
| HAW | 2.3 | 2.3 | 1.6 | 1.6 | 1.5 | 1.5 | 1.6 | 1.6 | 1.7 | 1.6 | 1.5 | 1.6 | 1.5 | 1.6 | 1.6 | 15.6 |
| MELB | 1.9 | 2.0 | 1.0 | 1.0 | 1.0 | 0.9 | 1.1 | 1.2 | 1.2 | 1.3 | 1.2 | 1.4 | 1.4 | 1.4 | 1.5 | 16.6 |
| NTH | 2.4 | 2.3 | 1.3 | 1.1 | 1.1 | 1.1 | 1.2 | 1.3 | 1.1 | 1.2 | 1.3 | 1.3 | 1.3 | 1.2 | 1.3 | 14.1 |
| PORT | 0.9 | 1.0 | 1.2 | 1.4 | 1.4 | 1.8 | 1.7 | 1.9 | 1.9 | 1.9 | 1.7 | 1.6 | 1.7 | 1.6 | 1.6 | 14.7 |
| RICH | 2.2 | 2.1 | 1.1 | 1.1 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 12.1 |
| STK | 2.2 | 2.1 | 0.9 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.9 | 0.9 | 0.8 | 0.8 | 0.9 | 0.9 | 10.8 |
| SYD | 2.5 | 2.4 | 1.0 | 0.9 | 0.9 | 1.0 | 1.0 | 0.9 | 0.9 | 0.9 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 12.7 |
| WESTC | 1.4 | 1.3 | 1.3 | 1.2 | 1.1 | 1.6 | 1.7 | 1.8 | 1.7 | 1.7 | 1.8 | 1.9 | 1.8 | 1.7 | 1.8 | 18.4 |
| WsB | 2.1 | 2.0 | 1.4 | 1.4 | 1.3 | 1.2 | 1.3 | 1.3 | 1.3 | 1.4 | 1.3 | 1.3 | 1.3 | 1.3 | 1.3 | 15.0 |

Table 3: The estimated annual expected number of available sons for each club from 1986 to 2015. Shaded cells indicate years where there were one (light shaded) or two (dark shaded) Father-Son selections for a club. The vertical lines delineate periods of substantial rule changes (Table 2; Tuck, 2015 Table 1a). The estimated total expected number of sons over the 10-year period 2016 to 2025 is also shown.

## 5. CONCLUSIONS

- A demographic model that estimates the annual number of expected sons by AFL club was largely able to explain differences in observed Father-Son selections
- The model can also explain the lower number of selections for the non-VFL aligned clubs
- Projections of expected sons by club showed that some clubs will have substantially more expected sons over the next 10 years than other clubs
- The model could be used as a decision making tool if further changes to the Father-Son rule are required for competitive balance

| | Since 1986 (VFL-aligned) | Since 1987 (WESTC) | Since 1991 (ADEL) | Since 1995 (FREM) | Since 1997 (PORT) | Since 2001 (SA/WA changes) |
|---|---|---|---|---|---|---|
| ADEL | | | 24.0 | 22.0 | 21.3 | 19.9 |
| BRIS | | 34.3 | 34.1 | 33.8 | 31.4 | 22.1 |
| CARL | 70.1 | 65.3 | 46.6 | 34.7 | 30.6 | 23.1 |
| COLL | 75.6 | 70.3 | 49.9 | 36.2 | 31.7 | 23.0 |
| ESS | 71.6 | 67.2 | 49.0 | 36.6 | 32.1 | 23.1 |
| FITZ | 39.7 | 34.7 | 15.4 | 2.2 | | |
| FREM | | | | 22.3 | 20.9 | 17.9 |
| GEEL | 74.3 | 68.7 | 47.3 | 34.3 | 30.0 | 21.7 |
| HAW | 75.3 | 70.2 | 50.6 | 38.2 | 33.7 | 25.1 |
| MELB | 66.0 | 61.3 | 42.4 | 30.4 | 26.8 | 19.7 |
| NTH | 71.5 | 66.6 | 47.1 | 34.1 | 29.2 | 20.5 |
| PORT | | | | | 24.7 | 23.4 |
| RICH | 68.0 | 62.6 | 41.9 | 29.3 | 25.3 | 17.6 |
| STK | 63.2 | 58.4 | 39.8 | 27.7 | 23.5 | 15.3 |
| SYD | 71.7 | 66.0 | 43.6 | 30.3 | 25.8 | 17.2 |
| WESTC | | 37.8 | 33.2 | 28.8 | 27.4 | 23.9 |
| WsB | 72.9 | 67.4 | 46.5 | 33.5 | 29.5 | 21.3 |

Table 4: The estimated expected number of available sons for each club summed across all years from a specified year until 2015. For example, as West Coast (WESTC) was established in 1987, their expected sons (shaded, 37.8) can be compared to other clubs by considering the 'Since 1987' column.

**Acknowledgements**

The author wishes to thank Col Hutchinson (VFL/AFL data) and numerous club historians and enthusiasts from the WAFL and SANFL (in particular Greg Wardell-Johnson and Mark Beswick).

**References**

AFL (2015). *Bidding System Review – Father-Son and Academy*. May 2015. Australian Football League production. 14pp.

Anderson, J. (2013). Geelong can look forward to a raft of potential father–son picks in the future. The Herald Sun 12 September 2013. Available at http://m.heraldsun.com.au/sport/afl/geelong-has-a-raft-of-potential-father-son-picks-in-the-future/ story-fni5f6yf-1226717999156?nk=9963d4c596d95f4a79135716ea4829be

Australian Bureau of Statistics (2015a). Births, Australia, 2014. Reference data 33010do001_2014. Canberra: ABS.

Australian Bureau of Statistics (2015b). Births, Australia, 2014. Reference data 33010do009_2014. Canberra: ABS.

Borland, J. (2006). Economic design and professional sporting competitions. Australian Economic Review, 39, 435–441.

Hillier, K. (2006). Like father like son: Legends of AFL football. Sydney: Macmillan Publishers Australia.

Macdonald, R. D., & Booth, R. (2007). Around the grounds: A comparative analysis of football in Australia. In B. Stewart (Ed.). The games are not the same: The political economy of football in Australia (pp. 236–331). Carlton, Victoria: Melbourne University Press.

Niall, J. (2012). The wins of the father. The Age 13 May 2012. Available at http://www.theage.com.au/afl/afl-news/the-wins-of-the-father-20120512-1yjxt.html

Quayle, E. (2014). Adelaide knocked back on father–son draftees, eight years after missing Bryce Gibbs. The Age 3 April 2014. Available at http://www.theage.com.au/afl/afl-news/adelaide-knocked-back-on-fatherson-draftees-eight-years-after-missing-bryce-gibbs-20140402-zqptj.html#ixzz2xt8gfz4F

Rucci, M. (2009). Father–son rule misses mark for Williams. The Advertiser 29 May 2009. Available at http://www.adelaidenow.com.au/sport/afl/ father-son-rule-misses-mark-for-williams/story-e6freco3-1225717897356

Tuck, G.N. (2015). From father to son: A review and demographic analysis of the Australian Football League's Father–Son rule. Sport Management Review, 18, 596-608.

Tuck, G.N., Macdonald, R.D. & Whitten, A.R. (2015). Management reference points for sporting leagues: simulating league expansion and the effect of alternative player drafting regulations. In Andreff, W. (Ed.). *Disequilibrium Sports Economics: Competitive Imbalance and Budget Constraints* (pp. 50-103) Cheltenham: Edward Elgar Publishing.

# UNDERSTANDING THE IMPACT OF DEMAND FOR TALENT ON THE OBSERVABLE PERFORMANCE OF INDIVIDUALS

Paul J. Bracewell [a,b,c], Evan Blackie [a], Peter Blain [b], Chris Boys [b]

[a] *DOT loves data*
[b] *CoderCred*
[c] *Corresponding author: paul@dotlovesdata.com*

## Abstract
The depth and quality of individuals within an available talent pool is found to be a function of demand. This is demonstrated by comparing globally-scaled individual performance metrics with measures of demand.

In order to find a substantial population base with a wide, deep reaching audience of varied ability and quantifiable performance metrics, computer programming ability is evaluated across a variety of languages. Public code repositories, such as GitHub, are accessed and the code quality assessed algorithmically.

The underlying family of probability distributions for measures of computer programming ability is roughly gamma for each programming language. Importantly, the shape parameters for each gamma distribution are positively correlated with the number of job advertisements listed.

Given the nature of this relationship, as the number of job advertisements for a programming language increases, the distribution tends to become less right skewed. From a practical perspective, the upper tails of the distribution are most telling when looking to build an elite team, whether this be for rugby, cricket or a group of developers within a business environment. The presence of fatter upper tails indicates a greater prevalence of high performing talent within a neighbourhood of acceptable performance. In instances where there is lower demand, the tails are thinner indicating a greater gap in ability between high performers and those of lesser quality.

These findings have longer term ramifications for understanding the quality of representative teams for sports which are losing player numbers, such as rugby in New Zealand. Alternatively, the underlying metrics could potentially be used to forecast elite performance or monitor throughput within academy environments.

**Keywords**: Gamma Distribution, Performance Monitoring

## 1. INTRODUCTION
The depth and quality of individuals within an available talent pool is found to be a function of demand. This is demonstrated by comparing globally-scaled individual performance metrics with measures of demand. While the research framework is applied to coding ability in this context, the approach could be expanded to look at historical participation rates within sport and team ratings.

## 2. BRIEF REVIEW OF PROBABILITY DISTRIBUTIONS AND RATINGS
There is no shortage of academic research that examines the performance of individuals, groups and teams within a wide variety of contexts (e.g. Clarke *et. al.*, 2011; Cook *et. al*., 1988; Croucher, 2000; Di Salvo *et. al*., 2010; Gerber *et. al*., 2009; Kimber *et. al*., 1993; Lemmer, 2004).

Concepts exploring latent factors, such as intelligence, are well researched (e.g. Zhang *et. al*., 2011; Moon *et. al*., 2005; Silvia *et. al.*, 2008; Cummings *et. al.* 2005). Substantial research has also been invested into the underlying probability distributions for performance metrics and individual ratings for a variety of sports (e.g. Bracewell *et. al.*, 2009; Bukiet *et. al*., 2006; Damodaran, 2006; Kimber *et. al*., 1993; Koulis *et. al*., 2014). However, individual privacy restricts open access to individual performance data, particularly about on-the-job performance. For this reason, elite sport data, which is highly prevalent on the internet (e.g. www.espncricinfo.com, www.optasports.com, www.championdata.com, www.espnscrum.com) has been used for a range of analytical exercises regarding quantifying performance and ability.

Sports research has also investigated the relationship between the size of a talent pool and the expected performance. "Larger countries have a deeper pool of talented athletes and thus a greater chance of fielding medal winners (Bernard & Bruss, 2004, p. 413)". However, total gross domestic product (GDP) for a nation is a better predictor of national Olympic performance than size alone, which implies access to resources for investing in the talent development is critical to success (Bernard & Bruss, 2004). Another potential factor driving national performance is the internal competition within that nation. However, many competitions do not publish data from domestic or amateur competitions.

The impact of demand on the quality and depth of an elite talent pool has limited research in the sports sector. From an economic perspective, Rosen (1981) explored the earning potential for individuals in a variety of disciplines given quality, supply and demand constraints. Although similar, the emphasis in this research is to understand how demand shapes the depth and quality of a talent pool.

## 3. DATA SOURCING

In order to find a substantial population base with a wide, deep reaching coverage of individuals of varied ability and quantifiable performance metrics, computer programming ability is evaluated across a variety of languages. Public code repositories, such as GitHub, were accessed and the code quality assessed algorithmically. A random sample of 8264 individuals located in Australia and New Zealand was processed by an Australian-domiciled third party, Lumnify. Lumnify have developed several proprietary algorithms for quantifying code quality and code maintainability (Lumnify, 2015). These algorithms enable Lumnify to estimate the quality of developers across a wide range of programming languages. The underlying premise is that the observed outputs in these public repositories are a manifestation of ability.

To quantify demand, the number of Job Applications featuring various languages was obtained by searching the job website Seek (www.seek.co.nz, www.seek.com.au) on 16th December, 2015. The number of job advertisements within the Information, Communication and Technology (ICT) category and observed exclusively within New Zealand or Australia were recorded for each language of interest (see Table 1).

Data was also obtained from CoderCred (www.codercred.com) evaluating the performance of individuals on a range of pre-defined programming challenges. CoderCred measures programming performance by using a set of industry relevant, time bound programming challenges. The CoderCred Event Score mathematically combines the Accuracy, Timeliness and Difficulty from a series of challenges undertaken by a coder. The proprietary mathematical weightings that combine these three attributes into a meaningful predictive measure of performance are defined by a non-linear optimisation routine. The purpose of this data is to independently validate the Lumnify scores.

Given the commercially sensitive nature of the proprietary data acquired from Lumnify, the data are aggregated into performance-based bins (Table 1).

## 4. DATA PROCESSING

Independently, individual Lumnify scores were validated using CoderCred's scoring framework, which is constructed on the premise that performance on well-constructed, relevant tests is a strong predictor of future performance. The correlation between the Lumnify passively acquired scores and CoderCred's actively obtained challenge scores have a correlation of 0.78 (n=31).

The random sample was representative of the demand population, with a correlation of 0.74 observed between the observed sample counts and job demand (see Table 1).

Summary statistics, including the mean and standard deviation, were calculated from the aggregated data, using the midpoint of the score range. Skewness, g, was calculated using the adjusted Fisher-Pearson standardized moment coefficient using the sample standard deviation, s, and sample size, n, as follows:

$$g = \frac{n}{(n-1)(n-2)} \frac{\sum_{i=1}^{n}(x_i - \bar{x})}{s^3} \qquad (1)$$

The gamma distribution is widely used to model physical quantities that take positive values (Kimber, 1993) and is capable of representing a variety of distribution shapes (Smith, 1993). Given these reasons and the gamma distribution's ability to mimic the attributes of other distributions (Tough, 1999) makes it a useful basis for modelling programming behaviour by language.

From the sample mean, $\bar{x}$, and standard deviation, s, the shape, α, and scale parameters, β, for the gamma distribution were calculated as: $\alpha = \frac{\bar{x}^2}{s^2}$ and $\beta = \frac{s^2}{\bar{x}}$.

Overlaying the expected count upon the observed counts in the Figure 1 reveals that the gamma distribution appears to be a reasonable fit. To ensure that the counts in the under 35 and 36-45 bins do not visually dominate, a log scale is used to show the fit, particularly in the right-hand tail.

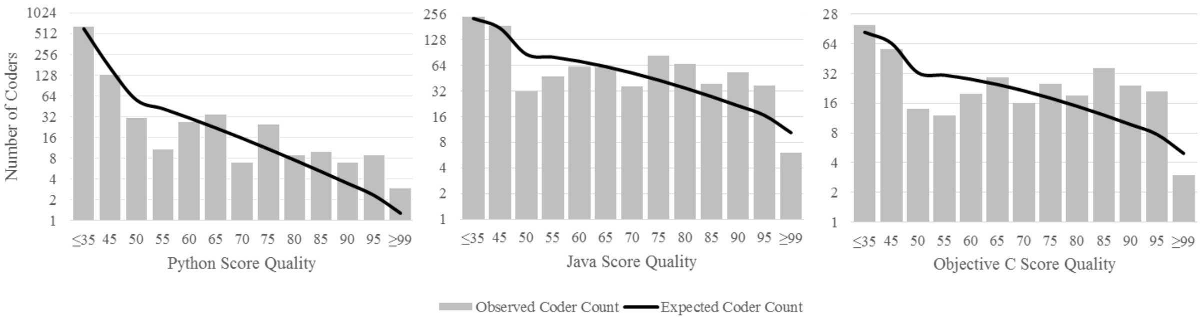



Figure 1: Histograms showing Coder Quality with Expected Distribution Overlaid.

| Quality Score | | Language | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | *C* | *C#* | *C+* | *Java* | *Javascript* | *Objective C* | *Perl* | *PHP* | *Python* | *Ruby* |
| *Number of Coders* | *<35* | 294 | 67 | 162 | 238 | 452 | 99 | 78 | 119 | 662 | 701 |
| | *45* | 94 | 60 | 107 | 187 | 279 | 56 | 3 | 79 | 130 | 52 |
| | *50* | 32 | 27 | 27 | 32 | 110 | 14 | 2 | 29 | 31 | 70 |
| | *55* | 27 | 22 | 24 | 47 | 280 | 12 | 0 | 48 | 11 | 13 |
| | *60* | 51 | 41 | 60 | 62 | 228 | 20 | 2 | 55 | 27 | 57 |
| | *65* | 37 | 41 | 40 | 61 | 124 | 29 | 2 | 39 | 35 | 46 |
| | *70* | 31 | 17 | 32 | 36 | 73 | 16 | 0 | 67 | 7 | 5 |
| | *75* | 23 | 49 | 35 | 83 | 133 | 25 | 1 | 58 | 25 | 24 |
| | *80* | 23 | 39 | 31 | 66 | 88 | 19 | 0 | 43 | 9 | 13 |
| | *85* | 29 | 41 | 36 | 39 | 114 | 36 | 0 | 95 | 10 | 6 |
| | *90* | 40 | 46 | 28 | 53 | 185 | 24 | 4 | 115 | 7 | 14 |
| | *95* | 37 | 31 | 23 | 37 | 152 | 21 | 1 | 60 | 9 | 8 |
| | *>99* | 4 | 3 | 2 | 6 | 26 | 3 | 0 | 11 | 3 | 0 |
| *Summary Statistics* | *Total* | 722 | 484 | 607 | 947 | 2244 | 374 | 93 | 818 | 966 | 1009 |
| | *Mean* | 46.8 | 61.1 | 51.5 | 52.4 | 56.0 | 54.5 | 29.3 | 63.5 | 32.5 | 33.3 |
| | *Stdev* | 24.5 | 22.1 | 22.8 | 22.9 | 23.0 | 24.6 | 17.3 | 23.3 | 17.2 | 18.0 |
| | *Skewness* | 0.50 | -0.36 | 0.15 | 0.13 | 0.02 | 0.00 | 2.64 | -0.51 | 1.75 | 1.49 |
| | *Demand NZ* | 86 | 192 | 38 | 174 | 247 | 43 | 9 | 58 | 66 | 45 |
| | *Demand AU* | 741 | 1195 | 362 | 1621 | 2037 | 306 | 232 | 810 | 539 | 381 |
| | *Demand* | 827 | 1387 | 400 | 1795 | 2284 | 349 | 241 | 868 | 605 | 426 |
| *Gamma Parameters* | *Alpha* | 3.6 | 7.6 | 5.1 | 5.1 | 5.9 | 4.9 | 2.9 | 7.5 | 3.5 | 3.4 |
| | *Beta* | 12.9 | 8.0 | 10.1 | 10.2 | 9.5 | 11.1 | 10.2 | 8.4 | 9.3 | 9.8 |
| *Observed Prop. ≥95* | | 5.7% | 7.0% | 4.1% | 4.5% | 7.9% | 6.4% | 1.1% | 8.7% | 1.2% | 0.8% |
| *Expected Prop. ≥95* | | 4.5% | 7.6% | 4.7% | 5.1% | 6.2% | 6.7% | 0.4% | 9.6% | 0.5% | 0.6% |

Table 1. Key Metrics for Assessing Coder Quality and Demand in Australia and New Zealand

## 5. SUMMARISING PERFORMANCE WITH PROBABILITY DISTRIBUTIONS

Comparing the cumulative observed and expected percentages per language using scatterplots suggested that the gamma distribution was a reasonable approximation. However, a goodness-of-fit test was applied to confirm that the data approximately followed a gamma distribution. Given the aggregated nature of the data, a Pearson's chi-square test was used to measure goodness-of-fit. However, these results indicated the null hypothesis, that the data fits a gamma distribution, cannot be accepted at a 5% level of significance for any of the languages. To test this hypothesis more rigorously, more granular data is required.

If the gamma distribution is considered an inappropriate fit, measuring the skewness for each distribution has the effect of summarising the underlying behaviour of interest. That is, the lower the demand, the more the distribution becomes right skewed. The implication is that for jobs with lesser demand, the thinner tails observed in a right skewed distribution imply that there is likely to be a disproportionally smaller talent pool. This effect is observed below in Figure 2, where the natural log of demand is plotted against skewness.

However, the shape of the distribution may have no bearing on the proportion of individuals capable of performing at a certain level, due to the influence of the underlying scale and location of the distribution describing code quality. To test this impact, demand is compared with the observed proportion of individuals with a code quality score of 95 and above. Figure 3 below shows that as the natural log of demand increases, so does the percentage of individuals scoring 95 or above for overall code quality.

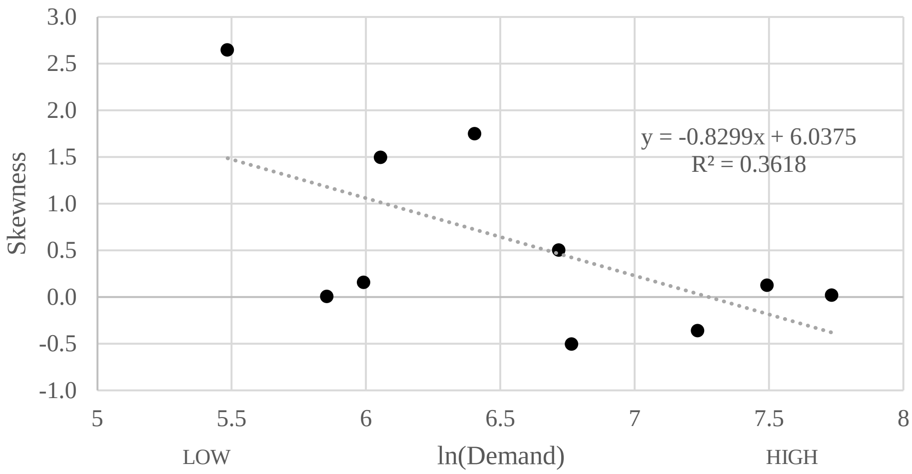



Figure 2. Scatterplot with trend line showing the negative relationship between the natural log of job demand in different languages and the skewness of the observed code quality distribution.

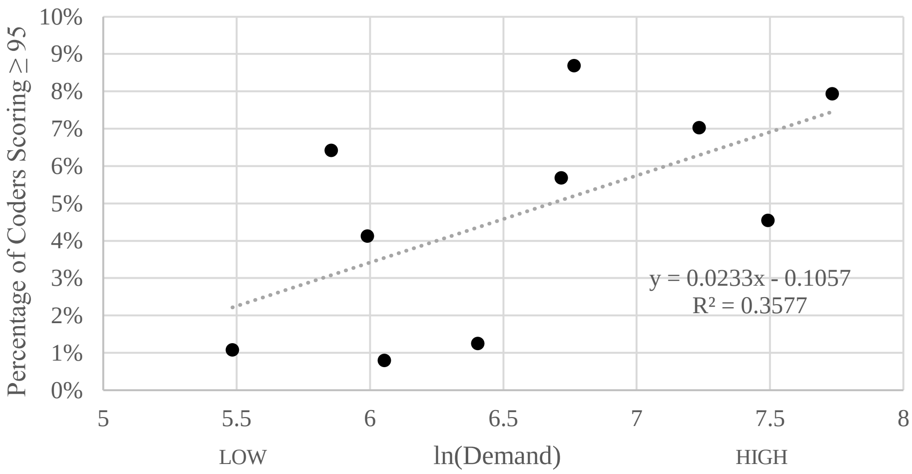



Figure 3. Scatterplot with trend line showing the positive relationship between the natural log of job demand in different languages and the percentage of coders scoring 95 or above.

Although statistically the fit of the gamma distribution was rejected, it is useful for explaining the behaviours in the tail. Using the fitted gamma distribution for each language, the correlation between the expected percentage of observations scoring 95 and above and the corresponding observed number of observations is 0.96. More importantly, the linear relationship to predict the expected percentage from the observed percentage had an intercept of 0 and slope of 0.99. Furthermore, the correlation between the observed skewness for each distribution and the inverse square root of the corresponding shape parameters α is 0.92. As skewness for the gamma distribution is $2\alpha^{-0.5}$, this result is not unexpected.

These last two points suggests that it is worthwhile to explore the probability distribution for coder quality further using actively acquired data where greater controls regarding data collection are present. This data will be obtained from CoderCred in the future.

**6. DISCUSSION AND CONCLUSION**

As the number of job advertisements for a programming language increases, the distribution of coder quality tends to be less right-skewed. The assumption is that because right-skewed distributions have thinner tails the relative depth of coder quality is lower in programming languages where there is less demand. To validate this assumption, the relative thickness of the right-hand tail of the distribution was assessed by comparing the observed percentage of individuals with code quality of 95 or above. This confirmed that there is a relationship between demand and the relative depth of the talent pool. Whether in the programming community the depth of talent pool is driven by competition, knowledge sharing or availability of resources, this phenomenon has wider implications for a range of business decisions. These decisions will range from hiring policies, training and development, adoption of technologies and a greater appreciation for the true cost of maintainability.

From a practical perspective, the upper tails of the distribution are most telling when looking to build an elite team, whether this be for rugby, cricket or a group of developers within a business environment. The presence of fatter upper tails indicates a greater prevalence of high performing talent within a neighbourhood of acceptable performance. In instances where there is lower demand, the tails are thinner indicating a greater gap in ability between high performers and those of lesser quality.

Further work is required to understand if this behaviour observed in the programming community is mirrored in elite team sport. Understanding this relationship would assist sporting organisations in demonstrating the longer term return on investment for community development initiatives. These findings have longer term ramifications for understanding the quality of representative teams for sports which are losing player numbers, such as rugby in New Zealand (Mathers, 2014; Napier, 2015; Robson, 2014). Alternatively, the underlying metrics could potentially be used to forecast elite performance or monitor throughput within academy environments.

**Acknowledgements**

We thank Milo Hyben of Lumnify for access to data extracted from the Lumnify scoring system and Hamish McEwen of Sports NZ for access to participation trends in NZ sport.

**References**

Bernard, A. B. & Busse, M.R. (2004). Who wins the Olympic Games: Economic resources and medal totals. *The Review of Economics and Statistics*, **86**(1), 413-417.

Bracewell, P.J. (2003). Monitoring meaningful rugby ratings. *Journal of Sports Sciences*, 21, 611-620.

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

Clarke, S. R. (2011). Rating non-elite tennis players using team doubles competition results. *Journal of the Operational Research Society*, **62**(7), 1385-1390.

Bukiet, B., & Ovens, M. (2006). A mathematical modelling approach to one-day cricket batting orders. *Journal of sports science & medicine*, **5**(4), 495.

Cook, W. D., Golan, I., & Kress, M. (1988). Heuristics for ranking players in a round robin tournament. *Computers & operations research*, **15**(2), 135-144.

Damodaran, U. (2006). Back Issues. *Journal of Sports Science and Medicine*, **5**, 503-508.

Di Salvo, V., Baron, R., González-Haro, C., Gormasz, C., Pigozzi, F., & Bachl, N. (2010). Sprinting analysis of elite soccer players during European Champions League and UEFA Cup matches. *Journal of Sports Sciences*, **28**(14), 1489-1494.

Farinaz, F. (2009). Was Bradman Denied His Prime? *Journal of Quantitative Analysis in Sports*, **5**(4), 1-26.

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

Kimber, A. C. & Hansford, A. R. (1993). A Statistical Analysis of Batting in Cricket. *Journal of the Royal Statistical Society*. Series A, **156**(3), 443–455.

Koulis, T., Muthukumarana, S., & Briercliffe, C. D. (2014). A Bayesian stochastic model for batting performance evaluation in one-day cricket. *Journal of Quantitative Analysis in Sports*, **10**(1), 1-13.

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

Mark Cummings, E., Keller, P. S., & Davies, P. T. (2005). Towards a family process model of maternal and paternal depressive symptoms: Exploring multiple relations with child and family functioning. *Journal of Child Psychology and Psychiatry*, **46**(5), 479-489.

Mathers, J. (2014, July 13). Battered and bruised – kids are shunning rugby. *NZ Herald.* Retrieved from www.nzherald.co.nz.

Moon, S. M., & Illingworth, A. J. (2005). Exploring the dynamic nature of procrastination: A latent growth curve analysis of academic procrastination. *Personality and Individual Differences*, **38**(2), 297-309.

Napier, L. (2015, June 28). Auckland club rugby facing crisis as player numbers take significant hit. *Stuff.* Retrieved from www.stuff.co.nz.

Robsen, T. (2014, May 8). Wellington club rugby suffers drop in playing numbers. *Dominion Post.* Retrieved from www.stuff.co.nz/dominion-post.

Rosen, S. (1981). The Economics of Superstars. *The American Economic Review*, **71**(5), 845-858.

Segall, D. O. (2001). General Ability Measurement: An Application of Multidimensional Item Response Theory. *Psychometrika*, 66(1), 79-97.

Silvia, P. J. (2008). Another look at creativity and intelligence: Exploring higher-order models and probable confounds. *Personality and Individual differences*, **44**(4), 1012-1021.

Smith, P.J. (1993). *Into Statistics*. Melbourne: Nelson.

Tough, R. J. A., & Ward, K. D. (1999). The correlation properties of gamma and other non-Gaussian processes generated by memoryless nonlinear transformation. *Journal of Physics D: Applied Physics*, **32**(23), 3075.

Zhang, Y., Zheng, Z., & Lyu, M. R. (2011, October). Exploring latent features for memory-based QoS prediction in cloud computing. In *Reliable Distributed Systems (SRDS), 2011 30th IEEE Symposium on* (pp. 1-10). IEEE.

# THE ROLE OF CLASSIFICATION IN THE DEVELOPMENT OF WEARABLE COACHING DEVICES

D P L Hunt Author [a,d] , Dave Parry [a,]

[a] *AUT University*

[d] *Corresponding author: dplhunt@gmail.com*

## Abstract

Current sensor, battery power and computing technologies allow for the feasible development of wearable coaching devices that can provide active, real-time feedback to athletes as they train. Single purpose devices that, for example, measure grip pressure for golf or pedalling cadence in cycling have been developed and are being used successfully. More general, quantification, systems that can identify (classify) and quantify durative activities with similar characteristics such as walking, running and cycling have been developed for smart phones and smart watches. However, systems capable of identifying multiple activities and applying different active, real-time feedback with differing requirements such as those needed for a backhand versus forehand shot in tennis have yet to be developed. One of the barriers to the development of multi-activity feedback systems is the complexity related to classifying those activities in real-time. Recent research has produced reliable classifiers for a number of durative activities but little work has been done on successful punctual activity classifiers within sports. Punctual activities are short, non-cyclic activities such as getting on a horse, turning a door handle or picking up a cup while durative activities occur over a longer time period and generally have a cyclic or rhythmic nature such as walking, running, cycling, rowing and swimming. The sensor signals associated with very short, punctual activities can be difficult to differentiate from the background signal. This work discusses the role of activity classification within real-time feedback systems, identifies a gap in the research related to punctual activity classifiers and briefly reports on ongoing research to address this gap using recurrent neural networks in the form of Echo State Networks.

**Keywords**: eCoaching, Punctual Activity Classification, Activity Classification, Wearable Sensors, Echo State Networks, Reservoir Computing

## 1. INTRODUCTION

Providing appropriate feedback is one of the fundamental roles of a coach and so should also be fundamental to embedded and/or wearable electronic coaching devices that provide real-time feedback on technique. Sports coaching feedback can take a number of forms with the aim of improving physical conditioning (strength, endurance and rehabilitation training), psychological performance (confidence, goal setting and strategising) and technique (technical and physical skills) (Cote & Gilbert, 2009; Strand, Benson, Buck, McGill, & Smith, 2014). This work concentrates wholly on providing feedback on technique and any associated technology.

A number of embedded and wearable devices have been developed to assist with providing some of this feedback. Most of these devices have been single purpose devices that measure, quantify or provide feedback on a single activity. More general purpose systems and devices capable of quantifying a variety of activities have been developed for smart phones, smart watches and fitness monitoring (e.g. FitBit).

With single purpose/single activity devices activity recognition (classification) is generally much simpler and sometimes trivial. The more generalised devices that quantify a number of different activities need more sophisticated activity classifiers. There are now generally available classifiers that are quite reliable for a small group of activities. These include such activities as walking, running and cycling, for example. On close inspection it will be noted that this group of activities has common characteristics, they tend to be longer activities with a rhythmic component to them and may be classed as Durative activities (D. P. Hunt & Parry, 2015). Some short, non-rhythmic activities (Punctual activities) such as swinging a baseball bat (e.g. Zepp smart Baseball bat) or serving in tennis (e.g. Zepp Tennis kit) do have embedded and/or wearable devices associated with them but these devices are single purpose and their associated activity classifiers are generally simplistic.

The accepted techniques used to classify durative activities such as walking, cycling and running utilise the rhythmic nature of these activities and often use features from the frequency domain in their classifiers. This can produce reliable results (Takacs et al., 2014), but these classification algorithms and in particular the frequency domain features that they rely on take some time to adjust at the start and end of an activity. This is generally not an issue for these durative activities because missing the first and last couple of seconds of running is not a big deal when running for ten minutes but it does mean that these techniques are much less reliable when used for punctual activities that may occur in less than a second or two. This highlights that there is a gap in the research and a need for classifiers and associated features that are capable of reliably recognising punctual activities.

## 2. WHAT USE IS AN ACTIVITY CLASSIFIER?

Activity classifiers have a wide range of uses, not only within a sports coaching context but also within a much wider context. Activity Classification Systems (ACS) are a key component of any system that purports to adapt to what a human being is doing at a particular time. ACS are also useful in an auditing context, e.g. Did the nurse unscrew the cap of pill bottle?; Did the patient do their physiotherapy exercises?; Did the automotive assembly worker fit the part using the required technique?; Did the trainee surgeon use the scalpel with the recommended pressure? Did the athlete do the requested number of squat exercises?

Within coaching, ACS are used to identify the current activity so that it can be quantified and/or recorded, as an adjunct to make the device "smarter" and more usable and/or to identify the current or next activity in order to provide appropriate feedback on that activity. In wearable coaching systems that provide real-time feedback they are especially necessary to ensure that the feedback is only provided when needed. The wearable riding coach proposed in Schliebs, Kasabov, Parry, and Hunt (2013) requires multiple ACS and these modules allow the overall system to identify the current activity, to help predict the possible next activity and to ensure that any coaching feedback is only provided when needed. For example, an equestrian athlete may be wearing the riding coach but not (yet) riding and so it would be pointless to provide feedback until the athlete was actually riding.

In summary, an ACS is an essential component of almost every smart feedback system based on human motion and this is confirmed by Baca (2012, p. 2) where he states explicitly that "*One main basis of almost any intelligent feedback system or adaptive system is the successful recognition or classification of patterns underlying the human motion just performed*".

## 3. WHAT ARE DURATIVE AND PUNCTUAL ACTIVITIES?

Using a temporal lens human activities can be broadly grouped into three classes:

- **Durative** - activities that contain recurrent or cyclic data or data appearing to occur at intervals (e.g walking, running, standing still, rowing, cycling and grooming a horse) and which occur over a longer time.
- **Punctual** - short, specific activities that may not contain periodic data (e.g. Pick up a cup, get on a horse, bowl, kick or hit a ball).
- **Complex or Meta** - An activity that is composed of two or more Durative and/or Punctual activities (e.g. Cooking, Riding).

These definitions were proposed in D. P. L. Hunt, Parry, and Schliebs (2014). At this stage of research, classifying complex or meta activities is generally not attempted because of complexity.

The daily activities that were a focus of early on-body, sensor researchers such as walking, running, standing, ascending stairs, descending stairs and the like are durative activities as they relate to continuing, cyclic action. (Smith, 1999). The common classification approach for these types of activities is to use the Quantize, Model and Classify pipeline According to Preece et al. (2009, pp. 3), in almost all cases the sensor signal is divided into small time segment "windows" and these are considered sequentially (Quantise). Features are calculated/derived from each window so as to characterise the signal (Model), and then the features are used as input into a classification algorithm (Classify).

Activities can be said to have a start phase, a central phase and an end phase. With durative activities the central phase is repetitive and looking at walking, for example, it could be said that a person starts walking, then they are walking for some period and finally they will stop walking at some point. By and large, the features that are extracted from the sensor signals that are used to classify durative activities come from the longer, repetitive, central phase.

Punctual activities, such as picking up a cup, opening a door and mounting a horse are usually very short compared with durative activities. These punctual activities do not contain a repetitive cycle and their short duration means that using features from all three phases is important. In addition, many punctual activities are of variable length. This temporal variability means that using fixed width windows to extract features tends not to be very successful. In this area Junker, Lukowicz, and Troster (2004, pp. 1) states:

"*With a scheme that partitions e.g. the signal into segments of predefined length, it is very likely to miss the exact beginning and end of the relevant movements which is critical particularly for activities of short duration*".

Spectral type quantisation techniques that are often successfully used on cyclic durative activities are unlikely to be successful with punctual activities as, by definition, punctual activities are not cyclic in nature. This is supported by researchers such as Logan, Healey, Philipose, Tapia, and Intille (2007) who report much lower classification rates when classifying punctual activities using spectral techniques.

Something different is required to successfully classify punctual activities. Some researchers such as Amft, Lombriser, Stiefmeier, and Troster (2007); Junker, Amft, Lukowicz, and Troster (2008) keep the methods developed for durative activities but supplement the sensor signals with proximity sensors, sound sensors and other data to try to improve classification success. Our approach has been to modify the classification technique by using recurrent neural networks (Reservoir Computing techniques) in the form of Liquid State Machines (Schliebs & Hunt, 2012; Schliebs et al., 2013) and Echo State Networks (D. P. Hunt & Parry, 2015; D. P. L. Hunt et al., 2014). Reservoir computing methods have two advantages when classifying punctual activities. Firstly they do not require that the sensor signal be pre-segmented into fixed windows for feature extraction as they are capable of using the unsegmented raw sensor signals or features calculated on the fly such as running means. This means that no additional special techniques are required to cater for variable length activities. Potentially an activity that takes two seconds to complete can be classified alongside the same activity that took four seconds to complete. Secondly, the ability of Reservoir methods to utilise the temporal sequence of the sensor signals enables more reliable classification of punctual activities because the temporal sequence of these short activities tends to be highly descriptive. This is highlighted by the success that other researchers have had using these methods in classifiers with other temporally significant data such as natural language (Goodman & Ventura, 2006; Jaeger, Lukosevicius, Popovici, & Siewert, 2007).

## 4. BRINGING IT ALL TOGETHER

Technology has been used to monitor and provide information to coaches and athletes that assist with training in all three areas of coaching feedback described in the introduction. Within the area of technique feedback, technology has been used to record and compare current technique with previous performances, with opponents and/or with theoretical ideals. Traditionally the technology used has provided after-the-event feedback. Some well known examples of technology that provides after-the-event feedback are video products such as Silicon Coach and Expert Vision Analysis (Liebermann et al., 2002). Such systems often require active guidance from a coach, especially when used by novice athletes (Liebermann et al., 2002). Feedback can be informal advice "keep your elbow down", visual comparisons that show a more optimal technique alongside the recorded performance or "technology supplemented" where, for example, a more optimal throwing angle is calculated and superimposed over the top of the recorded performance.

With the availability of small, low powered MEMS inertial and other sensors, powerful but low power draw computer chips and advances in battery miniaturisation and performance, small, wearable devices have been developed that can record data for after-the-event feedback or, more recently, to record, analyse and provide real-time feedback as the activity is being performed.

Within the realm of technique coaching, after-the-event feedback is useful but requires an athlete to remember to modify her technique when next performing the particular action. Sometimes the athlete’s current technique has been habituated and so coaching requests to "remember to do it differently next time" are not always immediately effective. When the athlete is highly motivated and does remember then this form of feedback can sometimes take months to change behaviour. Reducing the time between performing an activity and receiving feedback about that activity can speed up changes to technique, especially if the feedback is consistently provided every time the athlete performs the activity.

Most off-body, video coaching products for technique training require specially set up environments in order to function correctly. This means that if an athlete trains outside of that special environment they will not get the feedback needed and so potentially the rate of change will be slowed. This requirement to train within a special environment can add cost to training and make training less convenient.

Wearable devices and devices embedded within sports equipment and clothing are able to be used in most environments and so have the potential to provide more consistent feedback on technique, however, most of these devices to date have concentrated on quantifying sports activity. Probably the best known are the wrist devices such as fitbit that record data (quantities) such as time spent being active, number of steps, distance walked/run, elevations encountered and sometime physiological data such as heart rate. Other, more specialised devices include the Zepp sensor that records swing statistics for Baseball, Softball, Golf and Tennis; Some, more specialised devices do provide limited feedback, such as audible warnings if heart rate is above a predefined level (physical conditioning), audible or haptic metronomes to synchronise cadence (technique) or pressure sensitive golf gloves that provide an audible warning when grip is too tight (technique), see figure 1 for an example.

Figure 1: A pressure sensing golf glove designed to provide real-time feedback on club grip pressure. *Copyright by H Hellström/Hippson.se (2008)* Retrieved from http://www.hippson.se/artikelarkivet/ryttartraning/din-nya-tranare-en-dator.htm?_qStr=doug, Reproduced with permission of Hippson.se.

Most off-body, video coaching products for technique training require specially set up environments in order to function correctly. This means that if an athlete trains outside of that special environment they will not get the feedback needed and so potentially the rate of change will be slowed. This requirement to train within a special environment can add cost to training and make training less convenient.

A number of these wearable devices that provide real-time feedback on technique exist but they are invariably single activity devices and, in addition, they also invariably require manual intervention to turn them on and sometimes more importantly, to turn them off so that the feedback is not disruptive when not performing the particular activity that they are designed for. Potential problems focussing on what is important may also arise if athletes wear two or more different feedback devices at the same time and they are triggered at the same or similar instants. Imagine the confusion for an athlete that is perhaps wearing five or six different devices and they all start lighting up like a Christmas tree. Of course, a good human coach knows not to provide too much feedback about different things at the same time but stand-alone feedback devices have no collective intelligence. One way to resolve the issues of multiple unconnected devices is to consolidate them into an integrated wearable riding coach and this is the ultimate end goal of the work that we are currently doing at AUT University.

## 4. WORKING TO REALISE THE CONCEPT

We identified a number of components that needed to be built before an integrated, wearable riding coach could be realised and a decision was made to work on the classifier components first. Within the requirements for the classifiers a gap in the research was identified for a reliable punctual activity classifier. Recent work has been done to develop a suitable classifier using an Echo State Network. Encouraging early results have been published including work on the most difficult area for activity classification, classifying unscripted real-world activities in actual riding situations. The most recent publication from this work was D. P. Hunt and Parry (2015) and future publications are planned.

## 5. CONCLUSIONS

It is now feasible to contemplate creating an integrated, electronic wearable coach that provides real-time feedback on technique for sports coaching and other situations. One of the key components of such a wearable coach would be a classifier or series of classifiers. A prior gap in the research associated with activity classifiers has been a reliable classifier for punctual activities. One possible solution to this gap is to use Reservoir Computing methods. Other potential classifier engines may also be successful, however, Reservoir methods in the form of an Echo State Network has been shown to work reliably with one particular punctual activity taken from Equestrian sport.

**Acknowledgements**

We wish to thank the School of Engineering, Computing and Mathematics at AUT University, Auckland, New Zealand for financing parts of this work.

**References**

Amft, O., Lombriser, C., Stiefmeier, T., & Troster, G. (2007). Recognition of user activity sequences using distributed event detection. In *Smart sensing and context,* (pp. 126–141). Springer.

Baca, A. (2012). Methods for recognition and classification of human motion patterns - A prerequisite for intelligent devices assisting in sports activities. In *MATHMOD 2012, proceedings,* (Vol. 14). Vienna, Austria.

Cote, J., & Gilbert, W. (2009). An Integrative Definition of Coaching Effectiveness and Expertise. *International Journal of Sports Science and Coaching*, 4(3), 307–323. doi: 10.1260/174795409789623892

Goodman, E., & Ventura, D. (2006). Spatiotemporal pattern recognition via liquid state machines. In *International joint conference on neural networks*, (p. 3848-3853). Vancouver, BC. doi: 10.1109/IJCNN.2006.246880

Hunt, D. P., & Parry, D. (2015). Using Echo State Networks to Classify Unscripted, Real-World Punctual Activity. In *Engineering Applications of Neural Networks* (pp. 369–378). Springer.

Hunt, D. P. L., Parry, D., & Schliebs, S. (2014). Exploring the applicability of reservoir methods for classifying punctual sports activities using on-body sensors. In *ACSC 2014*. Auckland, NZ: Australian Computer Society.

Jaeger, H., Lukosevicius, M., Popovici, D., & Siewert, U. (2007). Optimization and applications of echo state networks with leaky- integrator neurons. *Neural Networks*, 20(3), 335–352. doi: 10.1016/j.neunet.2007.04.016

Junker, H., Amft, O., Lukowicz, P., & Troster, G. (2008). Gesture spotting with body-worn inertial sensors to detect user activities. *Pattern Recognition*, 41(6), 2010–2024. doi: 10.1016/j.patcog.2007.11.016

Junker, H., Lukowicz, P., & Troster, G. (2004). Continuous recognition of arm activities with body-worn inertial sensors. In *Wearable computers, 2004 8th IEEE international symposium on* (Vol. 1, pp. 188–189).

Liebermann, D. G., Katz, L., Hughes, M. D., Bartlett, R. M., McClements, J., & Franks, I. M. (2002). Advances in the application of information technology to sport performance. *Journal of sports sciences*, 20(10), 755–769.

Logan, B., Healey, J., Philipose, M., Tapia, E. M., & Intille, S. (2007). A Long-Term Evaluation of Sensing Modalities for Activity Recognition. In J. Krumm, G. D. Abowd, A. Seneviratne, & T. Strang (Eds.), *UbiComp 2007: Ubiquitous Computing* (pp. 483–500). Springer Berlin Heidelberg.

Preece, S. J., Goulermas, J. Y., Kenney, L. P. J., Howard, D., Meijer, K., & Crompton, R. (2009). Activity identification using body-mounted sensors – a review of classification techniques. *Physiological Measurement*, 30(4), 33.

Schliebs, S., & Hunt, D. P. L. (2012). Continuous classification of spatio-temporal data streams using liquid state machines. In *Iconip'12* (pp. 626–633). Qatar: Springer.

Schliebs, S., Kasabov, N., Parry, D., & Hunt, D. P. L. (2013). Towards a wearable coach: Classifying sports activities with reservoir computing. In C. J. L Iliadis H Papadopoulos (Ed.), *Engineering applications of neural networks* (Vol. 383, pp. 233–242). Springer.

Smith, C. S. (1999). Activities: States or Events? *Linguistics and Philosophy*, 22(5), 479–508. doi: 10.1023/A:1005439826271

Strand, B., Benson, D., Buck, R., McGill, W., & Smith, D. (2014). The characteristics of coaching expertise. *Virginia Journal*, 35(2), 4 – 7.

Takacs, J., Pollock, C. L., Guenther, J. R., Bahar, M., Napier, C., & Hunt, M. A. (2014). Validation of the Fitbit One activity monitor device during treadmill walking. *Journal of Science and Medicine in Sport*, 17(5), 496–500.

# TEAM PERFORMANCE AND CONFERENCE CALIBRATION

Clint A. Jowett, Samuel J. Rooney [a], Paul J. Bracewell [a,b]
[a] DOT loves data
[b] Corresponding author: paul@dotlovesdata.com

## Abstract
There are a vast range of methods used to create team ratings in sports. For example, Team Lodeings have been proven to give a good relative measure of team performance. Previous work has successfully applied these metrics to provincial rugby teams in New Zealand within the same division, but has not accounted for conference or division-type competition structures. In competitions with different conferences and divisions, such as the National Basketball Association (NBA) or New Zealand Rugby National Provincial Championship (NPC), there is usually varying ability of teams within each conference or division.

In this work, we use general linear models and smoothing to calibrate team ratings across groups. The NBA and National Football League (NFL) are used to highlight the limitations of Team Lodeings when comparing teams across divisions. By treating games played between teams within their own conference as one group, and those games between teams in different conferences as an interaction group, a general linear model is used to describe the relationship between these groups and to calibrate team ratings. For the NBA, the western conference has recently been stronger than the eastern, and our approach allows this unevenness to be incorporated into the competition-wide ratings.

We finally apply our methodology to the NPC and analyse the three divisions from 1976-2008. We are able to calibrate the ratings from teams between divisions using the general linear model approach. By comparing teams across divisions, we show the effect that professionalism has had on provincial rugby.

**Keywords**: Team ratings, calibration, general linear models

## 1. INTRODUCTION
Analytical techniques provide an excellent framework for comparing the relative merits between different teams. There are numerous publications describing the development and improvement of sport ratings systems. Stefani (2011) provides a detailed review of methods for officially recognised international sports rating systems and is an excellent resource for evaluating the strengths and weaknesses of various systems.

Bracewell, Forbes, Jowett, and Kitson (2009) introduced a method for quantifying the relative performances of teams, which used score ratios rather than scores or differences. This method, called Team Lodeings, enables meaningful comparisons of team ability, even if the teams have not played each other. However the framework does not account for differences between divisions. If there are no differences in the overall performance of teams between divisions, then intra-group Team Lodeings can be compared immediately. However, performance differences between divisions naturally arise in competitions like the New Zealand rugby domestic competition, NPC, or the English Premier League, where teams are promoted or relegated based on their season performance. In North American sports, differences between divisions occur because of asymmetric scheduling. This means that it is necessary to create a calibration framework for Team Lodeings in order to compare team ratings in multi-division competitions. This paper provides a framework for calibrating intra-group Team Lodeings so that overall ability differences between groups can be quantified.

## 2. METHODS
TEAM LODEINGS

We begin by providing an overview of the Team Lodeings ratings engine (Bracewell *et. al.*, 2009). The Team Lodeings rating, $L_T$, for a team $T$, is given by:

$$L_T = \frac{2\sum_{f}^{n_h} p_{T,f} \sum_{g}^{n_a} q_{T,g}}{n_h n_a} \qquad (1)$$

where $n_h$ and $n_a$ are the respective number of home and away games played by the $T^{th}$ team, $p_{T,F}$ is the ratio of victory for the $T^{th}$ team in the $f^{th}$ match at home, and $q_{T,g}$ is one minus the ratio of victory for the $T^{th}$ Team in the $g^{th}$ away match.

The ratio of victory for team $i$ against team $j$ is given by:

$$p_{i,j} = \frac{s_i}{s_i + s_j} \qquad (2)$$

where $s_i$ and $s_j$ are the normalized points scored by each team. The ratio of victory helps ensure ratings do not converge towards 0.5, as is the case in high scoring games such as basketball. Team Lodeings were originally developed for Rugby Union (Bracewell, *et. al.*, 2009), where the use of ratio of victory scores provided a well distributed sample of $\{q,p\}$. Here we use Rugby scores to standardize points scored, so that a variety of sports can be considered within this framework (with $\mu = 25$, $\sigma = 6.72$), so that the $\{q,p\}$ distributions are comparable. While we have chosen to use the ratings algorithm described above in this paper, an area for future research is the assessment of other algorithms.

Team Lodeings are calculated for a specified time frame, $t$, where $t \geq n_{h0} + n_{a0}$, with $n_{h0}$, $n_{a0} \geq 1$, so that Team Lodeings can be calculated. This allows for team performance to be dynamically calculated across a season so changes in performance can be quantified. This is useful for match prediction.

Higher Team Lodeings are associated with better performing teams, with $0 \leq L_T \leq 1$ ensured by the use of score ratios. Using score ratios leads to a well-defined $L_T$, whereas score difference can lead to unbounded $L_T$. Score ratios also lead to a fairer assessment of the performance of both teams in a given match. Consider two rugby results, 13-3, and 40-30. Both have a score difference of 10. However the second result reflects a closer game, which is reflected in the ratios of 0.81 and 0.57.

CALIBRATING TEAM LODEINGS ACROSS GROUPS

Our framework for calibrating Team Lodeings to account for division differences relies on having a suitable calibration group, or interaction group. This interaction group contains all matches played between teams from the different divisions we wish to compare. For example, for a competition with two divisions, we construct 3 distinct groups. Group A and Group B consist of all matches played exclusively between teams within division 1 and division 2 respectively, and Group C is the interaction group containing all matches between teams from division 1 and 2.

To create overall Team Lodeings, we use the following framework for calibration. First, we create calibration parameters:

$$P_{i,j} = \mu_A + \sigma_A A + \mu_B + \sigma_B B \qquad (3)$$

where $\mu_i$ is the mean adjustment to the Team Lodeings for division $i$, $\sigma_i$ is the multiplicative factor (standard deviation) to transform the Team Lodeings in division $i$, $A$ and $B$ represent the group, and $P_{i,j}$ is the score proportion in the match between teams $i$ and $j$ from groups A and B (Eq. (2)). Eq. (3) relates to matches within the interaction group only.

In practice, we fit these calibration parameters using the following general linear model:

$$P_{i,j}G_A + P_{j,i}G_B = \alpha_A G_A + \beta_A G_A L_i + \alpha_B G_B + \beta_B G_B L_j \qquad (4)$$

where $G_i$ are indicator variables for groups A or B. The model estimates are mapped to the intra-group Team Lodeings shift parameters by:

$$\alpha_A \to \mu_A \qquad (5)$$

$$\beta_A \to \sigma_A \qquad (6)$$

The calibrated Team Lodeings for team T in group A are then given by:

$$\bar{L}_T = \frac{L_T}{\sigma_A} - \mu_A \qquad (7)$$

where $L_T$ is the raw intra-group Team Lodeings, and $\bar{L}_T$ is the new calibrated Team Lodeings. For groups with relatively different levels of abilities, the recalibrated Team Lodeings will help factor for those differences that would otherwise be unaccounted for.

**3. RESULTS**

NATIONAL BASKETBALL ASSOCIATION: THE INFLUENCE OF DIVISIONS

We use the National Basketball Association (NBA) to provide a test of our framework for unbalanced leagues. The NBA is divided into Eastern and Western conferences with 15 teams in each. Within each conference there are three divisions. Each team plays 82 games, including 4 games against every other team in their own division. Teams face opponents from the other two divisions within the same conference either three or four times, and will play every team in the other conference twice. This asymmetrical structure means the strength of schedule will vary between teams, and calibrating team ratings will be important.

We consider all regular season NBA games played in the 2007-08 season. Given that the NBA is naturally split into two conferences, we define group 1 to include all matches within the Eastern conference, group 2 to include all matches within the Western conference, and the interaction group to consist of all matches between opponents from the Eastern and Western conferences.

To demonstrate the effect of conferences in the NBA, we calculate the Team Lodeings for each team from the Eastern conference group, Western conference group, and interaction group, and compare these with an Overall Team Lodeings measure, constructed without defining any conference sub-groupings. Figure 1 shows the correlation between grouped Team Lodeings and overall Team Lodeings for each NBA team at the end of the 2007-2008 season. There is a strong correlation for all groups (Eastern: $r$ = 0.96, $p$ < 0.0001, Western: $r$ = 0.95, $p$ < 0.0001, Interaction: $r$ = 0.87, $p$ < 0.0001). There is an equally strong correlation between the group Team Lodeings and final winning percentage (Eastern: $r$ = 0.90, $p$ < 0.0001, Western: $r$ = 0.90, $p$ < 0.0001, Interaction: $r$ = 0.89, $p$ < 0.0001). These correlations show that each individual group is able to produce Team Lodeings that are representative of the actual season.

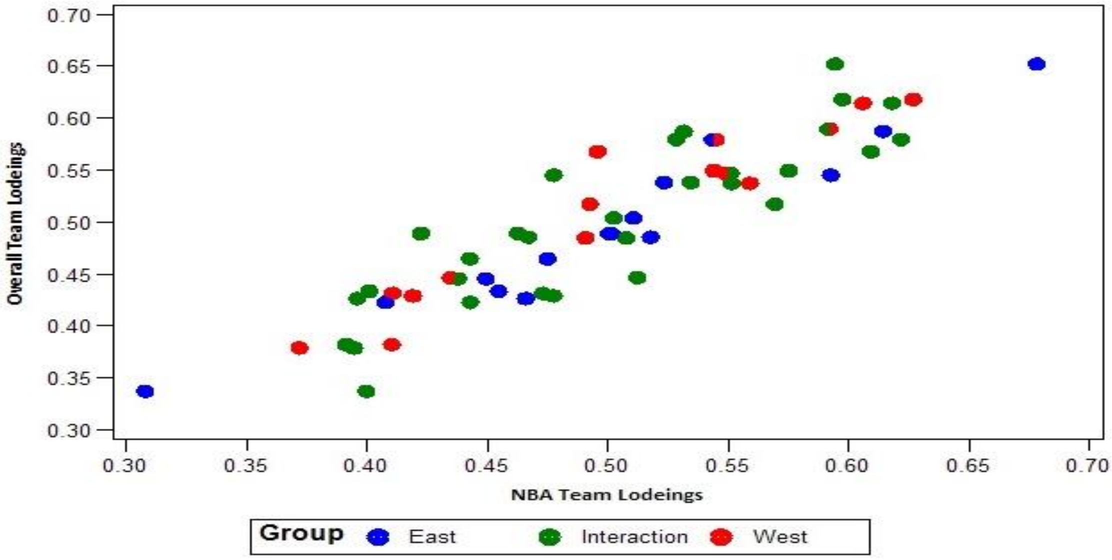



Figure 1: The relationship between Grouped NBA Team Lodeings and Overall NBA Team Lodeings for the 2007-2008 Season.

Although the Team Lodeings within each group are representative of the overall Team Lodeings, there are some important differences. The top rated team belongs to the Eastern Conference (Boston Celtics), where their own conference Team Lodeing is higher than that in the interaction group. Similarly the Eastern conference had the worst-rated team based on games played against their own conference (Miami Heat). Figure 1 immediately suggests the Eastern conference was less competitive than the Western conference in 2007-2008.

To quantify the differences between the groups, we use a general linear model with overall Team Lodeings as the dependent variable. Generalizing Eq. (4), we construct a model for the overall Team Lodeings of the form:

$$L_T = \sum_{\mathrm{i}} \mu_i G_i + \sigma_i G_i L_{G,T} \tag{8}$$

where $i$ is an index for the Eastern, Western and interaction groups, and $L_{G,T}$ is the intra-group Team Lodeings for team $T$, and $G_i$ is a group indicator variable. We find the model coefficients are given by:

$$\mu_E = 0.05, \mu_W = 0.02, \mu_I = 0.03 \tag{9}$$

$$\sigma_E = 0.86, \sigma_W = 0.97, \sigma_I = 0.93 \tag{10}$$

where $E$, $W$, and $I$ represent the Eastern, Western and Interaction group parameters respectively. The model estimates highlight the differences in ability between each group. The estimate for the Eastern conference shift ($\mu$) is highest showing this conference did not perform as well as the Western. Similarly, the multiplicative term ($\sigma$) is smallest for the Eastern Conference group showing there is a larger spread (unevenness in competition) in this conference.

NATIONAL FOOTBALL LEAGUE: A TEST OF CALIBRATION

The above analysis for the NBA showed how general linear modelling can be used to quantify the differences between divisions. However the correlation between each grouping's Team Lodeing and the overall Team Lodeing was very strong, which amounts to saying uncalibrated Team Lodeings work well when conferences are even. Here we consider the National Football League as an example to show how conference calibration performs across divisions with uneven performance.

The NFL has a similar conference system to the NBA, with the National Football Conference (NFC) and American Football Conference (AFC) containing 16 teams and 4 divisions. Teams play 16 games, playing every other team in their division twice, every team in one other division within their own conference once, and every team in one other division in the other conference once. Finally teams play two games against other teams in their own conference that finished in the same position in their own divisions as themselves in the previous season, not counting those in the division they were already scheduled to play.

In our study of the NFL, we consider all games played in the 2006-2007, 2007-2008, and 2008-2009 seasons (48 games per team). We find using the natural NFC and AFC groupings results in little difference between the Team Lodeings within each group, with a strong correlation between the overall Team Lodeings and group Team Lodeings ($r = 0.78$, $p < 0.0001$). To validate the calibrated Team Lodeings framework for groupings of differing ability, we engineer two groups based on ability. We ranked teams over their three season record and retained the top 8 teams in group 1, and assigned the bottom 8 teams to group 2. Figure 2 shows the correlation between overall Team Lodeings and the raw Intra-Division Ratings. Despite the differing ability between the two groups as shown in the Overall Team Lodeings, we find little distinction between the Intra-Division ratings for each group.

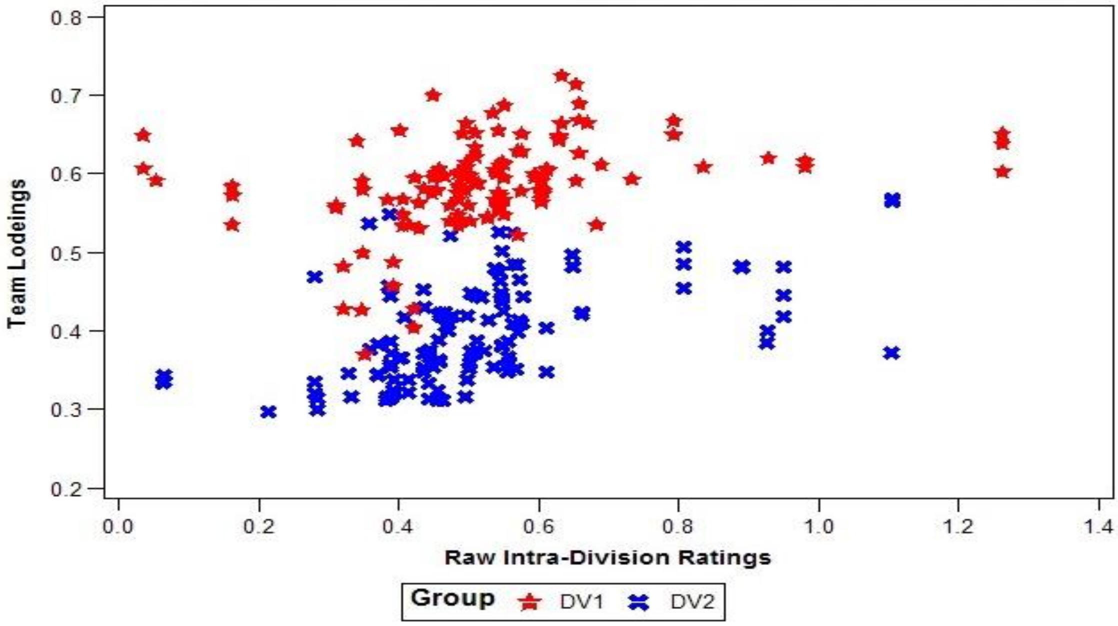



Figure 2: The relationship between Grouped NFL Team Lodeings and Overall NFL Team Lodeings for the 2006-2009 seasons.

In order to calibrate the intra-division ratings, we use the result proportion in a general linear model as in Eq (4). From games in the interaction group, we construct $P_{i,j}$ for each team from group 1 and 2, and use each team's raw intra-group Team Lodeings to fit the model given by Eq. (4). The model estimates are then used to calibrate the intra-division Lodeings via Eq (7). Figure 3 shows the correlation between the calibrated intra-division Team Lodeings and the Overall Team Lodeings. We now observe a clear separation between the two groups, and the calibrated ratings have a significant correlation with the overall ratings ($r = 0.68$, $p < 0.0001$).

NATIONAL PROVINCIAL CHAMPIONSHIP: APPLICATION OF THE CALIBRATION FRAMEWORK

We finally apply the calibrated Team Lodeings framework to the New Zealand National Provincial Rugby Championship (NPC). We calculate calibrated Team Lodeings for all teams from the inception of the NPC in 1976, until 2008. While the specific structure of the NPC has changed over the years, the general structure consists of three groups or divisions. In the inaugural competition, there was a top division consisting of provincial teams from throughout New Zealand, and a further two divisions based on North Island and South Island teams respectively. The format of the competition changed in 1986, when three truly national divisions were created. This corresponded to the emergence of clear differences in abilities of the three divisions.

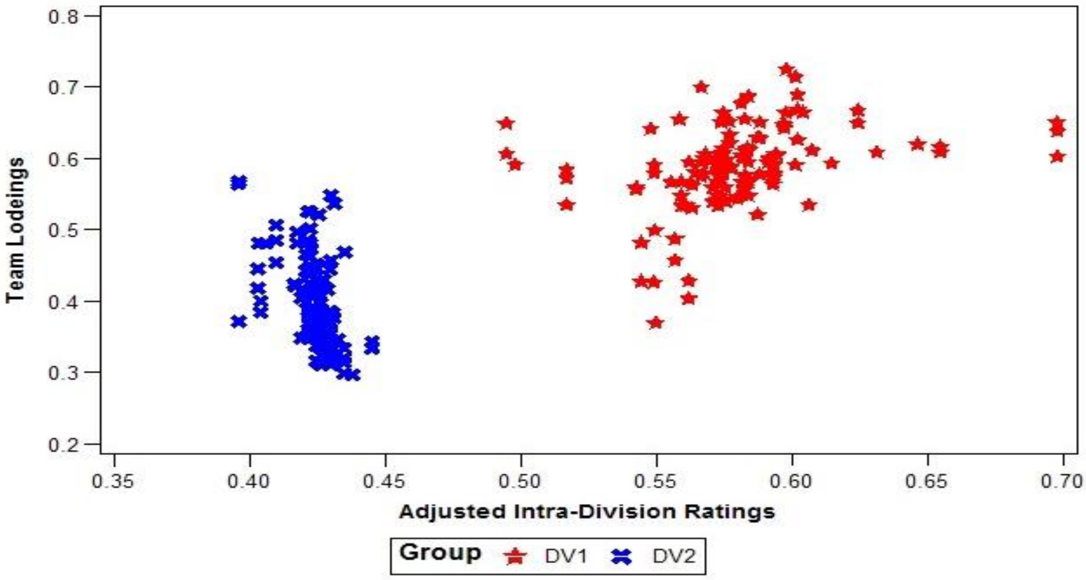



Figure 3: The relationship between the Calibrated Grouped NFL Team Lodeings and Overall NFL Team Lodeings for the 2006-2009 seasons.

In Figure 4 we present the calibrated ratings for three different teams – Auckland, Taranaki, and Wairarapa Bush. We choose to focus on these three sides for a number of reasons. Auckland is historically the most successful province, and has played in the top division in all seasons considered. Taranaki has been an established first division side since the mid 1990's, but prior to this had frequently moved between divisions. Finally, Wairarapa-Bush is one of the few sides to have played in all three divisions, and gives a good picture of how smaller rural-based unions have performed. The time series in Figure 4 can be interpreted by considering three distinct time periods:

*i: The birth of the NPC – 1976 to 1985.* The first 10 years of the competition are those where the three teams have the most even ratings. This aligns with the fact that division 2 and 3 were separated by geography, rather than ability. Wairarapa-Bush has its best period of ratings from 1981, corresponding to its promotion to the first division. Taranaki were relegated to the second division that same year. The general similarity between the ratings of Taranaki and Wairarapa-Bush shows the evenness between divisions 1 and 2 in the early years of the NPC. Auckland generally has a slightly higher rating until 1984, where their rating jumped after winning the division 1 competition.

*ii: Restructuring and shift towards professionalism – 1986 to 1995.* In this period, New Zealand provincial rugby was dominated by Auckland. They won 7 out of the 10 division 1 championships, including four in a row from 1987 – 1990, explaining their high Team Lodeings. Taranaki spent this period switching to and from divisions 1 and 2, where their promotion and relegation correlates with the oscillations in their Team Lodeings. Wairarapa-Bush were a regular mid-table second division team from 1988-1995. Over this period, an increase in the unevenness between divisions 1 and 2 is apparent. This corresponds to the restructuring of the competition to three national divisions driven by ability.

*iii: Professional era – 1996 onwards.* 1996 was the first year of truly professional rugby with the creation of the Super 12 competition. In this era, we see evidence of a strong drop off in the performance of rural provinces. Wairarapa-Bush were relegated to division 3 in 1998, where their rating dropped dramatically. In 2005 they won division 3, where we see their rating trending upward. However this time corresponds to the greatest gap between Wairarapa-Bush's rating, and those of Auckland and Taranaki, who were consistently performing division 1 sides. Wairarapa-Bush had a negative rating in 2004, highlighting the major differences that have developed between divisions in the NPC. This is a direct impact of professionalism where it is beneficial for the top players to go to the major regions of New Zealand.

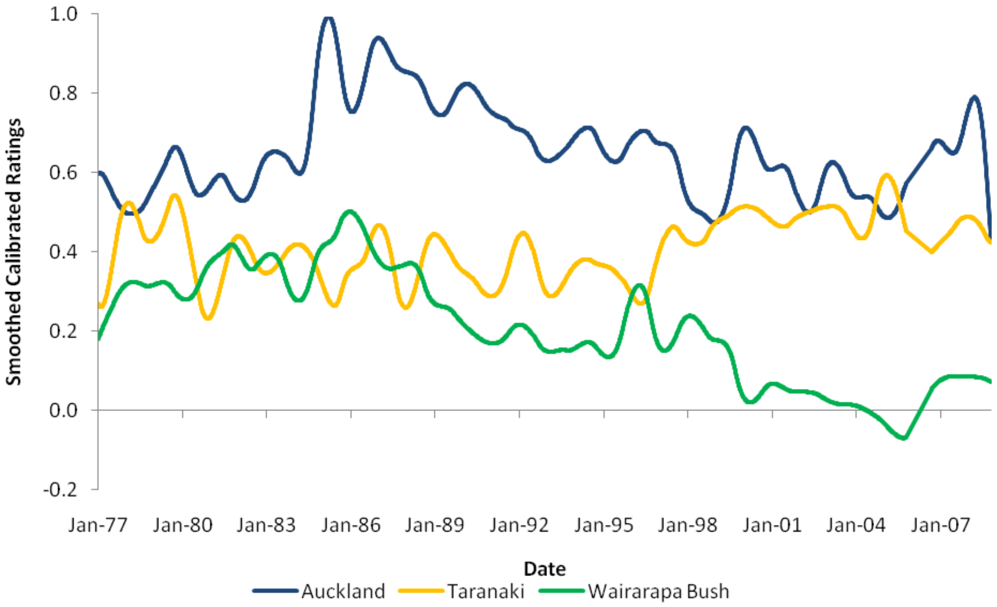



Figure 4: Calibrated NPC Team Lodeings for Auckland, Taranaki and Wairarapa Bush from 1976 – 2008.

## 4. CONCLUSIONS

We have presented a framework to extend Team Lodeings by calibrating intra-division Team Lodeings within a multi-division competition to account for differences between groups. We tested the approach on the NBA, highlighting strength of schedule effects, with the Western Conference Team Lodeings having the largest adjustment shift. By considering the NFL and NPC, we have shown that our calibrated Team Lodeings framework is able to compensate for ability differences between divisions. This is a generic framework, and can be applied to any multi-group competition in which intra-group Team Lodeings can be determined provided there is an interaction group to enable calibration.

## References


Bracewell, P.J., Forbes, D.G., Jowett, C.A., & Kitson, H.I.J. (2009). Determining the evenness of domestic sporting competition using a generic rating engine. *Journal of Quantitative Analysis in Sports* **5**(1).

Clarke, S.R. (2011). Rating non-elite tennis players using team doubles competition results. *Journal of the Operational Research Society*, **62** (7), 1385–1390.

Garland, S. J. (1997). *Fields of glory: 21 NPC years, 1976-1996*. Auckland, N.Z.: HarperCollins.

Knight, L. (1992). National Champs. In NZRFU (Ed.), *Clash of the Century. Centennial Programme 1892-1992*, 47-48. Wellington, N.Z.: NZRFU.

Leifer, E. M. (1995). *Making the majors: the transformation of team sports in America*. Cambridge, Mass.: Harvard University Press.

Marburger, D. R. (1997). *Stee-Rike Four! What's Wrong with the Business of Baseball?* London, England: Praeger Publishers.

McMillan, J. (1997). Rugby meets economics. New Zealand Economic Papers, **31**(1), 92-114.

Leitner, C., Zeileis, A., & Hornik, K. (2010). Forecasting sports tournaments by ratings of (prob)abilities: A comparison for the Euro 2008. *International Journal of Forecasting*, **26**(3), 471–481.

Stefani, R. (2011). The methodology of officially recognized international sports rating systems. *Journal of Quantitative Analysis in Sports*, **7**(4).

# APPLICATIONS OF COMPUTERISED COGNITIVE TESTING IN SPORT AND POTENTIAL IMPLICATIONS FOR CONCUSSION TESTING AND PLAYER AND COACHING DECISIONS

Adrian Schembri [a,b,c]

[a] *Cogstate Ltd.*
[b] *RMIT University*
[c] *Corresponding author:* aschembri@cogstate.com

## Abstract

For many years computerised cognitive testing has been integrated into elite sports worldwide. These tests have been used primarily for conducting baseline and after-injury assessments to aid in return to play decisions following a concussion. In this presentation, recent technological advances and improvements in the cognitive tests themselves will be discussed, particularly how these advances have enhanced the accessibility and usability of cognitive assessments for athletes and medical staff. The availability of online testing has also made it possible for baseline tests to be completed in an unsupervised setting (e.g., at home), which has increased accessibility for amateur athletes where team medical staff are not readily available to complete in person assessments. Despite several advantages, unsupervised testing poses a series of challenges in terms of data quality and comparability across repeat assessments given variations in the testing environment (and potential distractions) as well as differences in the level of understanding of the test requirements. Given this, the availability of training modules that ensure an athlete has sufficient understanding of the test requirements prior to the assessment is essential. Training modules also make cognitive tests more accessible for younger athletes, particularly children and adolescents who are playing contact sports where the risk of concussion remains. Aside from concussion testing, the potential applications of cognitive testing in the sporting context will be discussed, with a focus on how these tests can be used to facilitate player and coaching decisions both during training and in-play.

*Keywords*: Cognitive Testing, Concussion, Coaching Decisions

### 1. INTRODUCTION

Concussions occur in a broad range of contact sports worldwide, including American and Australian Rules football, rugby league, rugby union, boxing, hockey, lacrosse and basketball, and less commonly in sports such as soccer and netball. Given the frequency of falls, concussions are also frequent in sports such as horse riding and gymnastics. The incidence of concussion has been well known for several decades, with concussive assessments being undertaken both on the sidelines during matches and in the weeks subsequent to a match where a concussive event has occurred (or was suspected to have occurred).

According to the Sport Concussion Assessment Tool – 3rd Edition (SCAT3), “a concussion is a disturbance in brain function caused by a direct or indirect force to the head. It results in a variety of non-specific signs and/or symptoms and most often does not involve loss of consciousness. Concussion should be suspected in the presence of any one or more of the following: symptoms (e.g., headache), physical signs (e.g., unsteadiness), impaired brain function (e.g. confusion) or abnormal behaviour (e.g., change in personality).” (Concussion in Sport Group, 2013). There is general agreement that athletes should not be returned to play while injured and accordingly, many sports governing bodies have now developed concussion management programs with the aim of increasing awareness and improving management.

In recent years, sports governing bodies have become increasingly cautious in the manner that concussion is diagnosed and treated, with return to play decisions being more conservative as a result. This is particularly the case where an athlete has sustained multiple head injuries, given the association between multiple concussions and the increased risk of cumulative cognitive deterioration (i.e., the second impact syndrome) and mental illnesses such as depression (Makdissi et al., 2001; Makdissi, Darby, Maruff, Ugoni, Brukner, & McCrory, 2010). In rare cases, players have retired prematurely when multiple concussions of a severe nature has been sustained. In March this year, AFL player Justin Clarke, who had played 56 games over a four year period for the Brisbane Lions, retired at age 22 due to ongoing cognitive impairments endured following several serious concussions. Due to the risks associated with suffering further head trauma over the course of his career, it was recommended by Justin's treating team that he cease playing contact sport (Gleeson, 2016).

The emphasis on protecting players and minimising the incidence of concussions has had a flow-on effect to the assessment of symptoms and the manner in which the nature and severity of concussions are evaluated. There are a range of assessment tools available to assist club doctors at the elite level to assess concussion immediately following an incident during a match (Makdissi et al., 2001). Whilst some of these measures are conventional pencil and paper assessments, others are computerised neuropsychological assessments that enable a broad set of core cognitive domains to be evaluated in a short amount of time, usually within 10 to 15 minutes (Collie et al., 2003; Collie, Makdissi, Maruff, Bennell & McCrory, 2006). Decisions about the presence, magnitude and time course of concussion are limited when based on self-report symptoms alone. There is now strong evidence that the addition of objective assessments of CNS function such as cognition and balance improves the identification of concussion, with cognition returning to pre-injury levels after symptoms have resolved.

In this paper, current approaches to assessing cognition in concussion are considered, with an emphasis on computerised tools that can be utilised to evaluate cognition. Recent technological advances and improvements in cognitive tests are considered, particularly how these advances have enhanced the accessibility and usability of cognitive assessments for athletes and medical staff. The challenges associated with cognitive testing in elite and amateur sport are discussed, including methods that can be used to ensure a valid assessment of cognitive performance is being obtained. Implications for increasing accessibility to assessment tools at amateur levels are also considered, given the limited access to club doctors at the community sporting level.

## 2. ASSESSMENT OF CONCUSSIVE SYMPTOMS

Many assessment tools are available to evaluate concussive symptoms. Importantly, it is not recommended that a single assessment tool be used to make or exclude a concussion diagnosis, given the diagnosis of a concussion is a clinical judgment that should be made by a medical professional. In many cases, a medical practitioner will use a number of available tests in addition to a clinical interview to provide convergent evidence as to whether an athlete has sustained a concussion, and whether the athlete remains symptomatic and is ready to return to play (Makdissi et al., 2010).

### Sport Concussion Assessment Tool – 3rd Edition (SCAT3)

The SCAT3 is perhaps the most widely used assessment of concussion. This is a standardized assessment to evaluate concussion in athletes as young as 13 years of age. An alternative form, the Child SCAT3 is available for children aged 12 years and younger (Concussion in Sport Group, 2013). The SCAT3 assessment contains a general background which includes demographic information and personal details, as well as four questions that address the athlete's recent concussion history. A symptom evaluation checklist includes 22 common symptoms that an athlete may experience in the event of a concussion. The SCAT3 contains a cognitive assessment that evaluates cognitive domains such as orientation, memory (both immediate and delayed recall) and concentration. A neck, balance and coordination examination also form part of the SCAT3 assessment (Concussion in Sport Group, 2013). This tool is relatively simple to score and can be used to provide an indication of concussive symptoms immediately after a suspected head injury or in the hours, days or weeks following the event.

### Cogstate Brief Battery (CBB)

The SCAT3 assessment is often used in conjunction with a computerised cognitive screening assessment. Several brief computerised assessments are available, with some of the more common assessments being the Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT) and the Cogstate Brief Battery (CBB) (Covassin et al., 2009; Cromer et al., 2015). The CBB is a brief cognitive screening test battery that requires approximately 10-15 minutes to complete. The assessment consists of four tests that measure the speed and accuracy of psychomotor function (Detection task), attention (Identification task), learning (One Card Learning task) and working memory (One Back task). The validity of the CBB has been established in multiple studies of different healthy and clinical groups, including many studies on athletes of all ages (Collie et al., 2006; Eckner, Kutcher, & Richardson, 2010). The CBB has been shown to have high sensitivity to change, excellent specificity and sensitivity to cognitive impairment, high test/retest reliability as well as resistance to practice effects (Collie et al., 2003; Louey et al., 2014). Stimuli used in the CBB consist of cards from a traditional French card deck (refer to Figure 1), and therefore performance is not influenced by the language or cultural background of the athlete and therefore it can be administered validly in children, adults and older adults.


Figure 1: Screenshot of the Identification test in the Cogstate Brief Battery (CBB).

At the beginning of each CBB test, instructions are presented on the screen. This is followed by a demonstration in which athletes practice the test, thus providing an opportunity to become familiar with the rules. Once the demonstration is complete, the test begins. On each trial of each test, a single playing card stimulus is presented in the center of the screen. At the presentation of each playing card stimulus, individuals are required to respond either "YES" or "NO". The four tests contained within the CBB, the paradigm, main cognitive domain that each test measures and the primary outcome of each test are listed in the Table 1.

Table 1: Summary of tests in the Cogstate Brief Battery (CBB).

| **Test Name** | **Paradigm** | **Cognitive Domain** | **Primary Outcome** |
|---|---|---|---|
| Detection | Simple Reaction Time | Psychomotor Function | Speed of performance (mean of the log10 transformed reaction times for correct responses) |
| Identification | Choice Reaction Time | Attention | Speed of performance (mean of the log10 transformed reaction times for correct responses) |
| One Back | N-Back | Working Memory | Speed of performance (mean of the log10 transformed reaction times for correct responses) |
| One Card Learning | Pattern Separation | Visual Learning | Accuracy of performance (arcsine square root proportion correct) |

In leagues such as the AFL and National Rugby League (NRL), athletes complete a CBB assessment at the beginning of each season to provide a baseline level of cognitive function on each of the four tests in the battery. If the athlete sustains a head injury during the season, he or she then retakes the battery, whereby performance on each test is compared with the baseline assessment. Given after-injury performance is compared with baseline performance, it is critical that the athlete obtains a valid baseline assessment that accurately represents their level of cognitive function at the time of the test. However, given after-injury tests will typically be used to provide convergent evidence in return to play decisions, there is an incentive for athletes to feign poor performance on their baseline assessment, as an after-injury assessment will be less likely to identify cognitive decline if performance was poor at baseline. As such, the CBB has a series of in-built completion and data integrity checks that ensure each test has been completed and that performance was at the level expected for a healthy athlete who was not impaired by a lack of motivation, poor concentration or other potential distractions.

In analysing results of the CBB during after-injury assessments, meaningful decline is indicated if performance on one or more tests in the battery has declined by greater than 1.65 standard deviations when compared with the most recent baseline assessment (as measured by a Reliable Change Index). If cognitive decline has occurred, a retest is recommended and the CBB is typically repeated within one to seven days. It is not uncommon for the CBB to be repeated on several occasions in the week(s) following a concussion, refer to Figure 2 for an example.

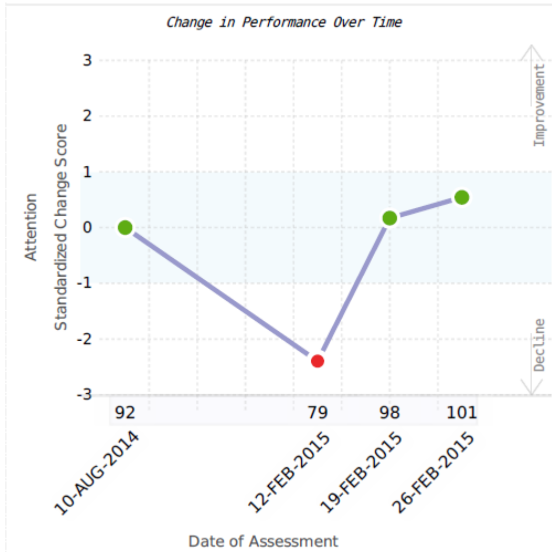



Figure 2: Example of a change score plot displaying change in cognitive performance over multiple assessments.

In addition to measuring cognitive decline relative to baseline, the CBB also provides a comparison to age-matched normative data on each test in the battery, refer to Figure 3. Performance on each test is represented as a standardized T-Score with a mean of 100 and a standard deviation of 10. If performance is more than two standard deviations below the normative mean for the athlete's age (e.g., <80), cognition is considered to be impaired, in which case a retest is recommended.

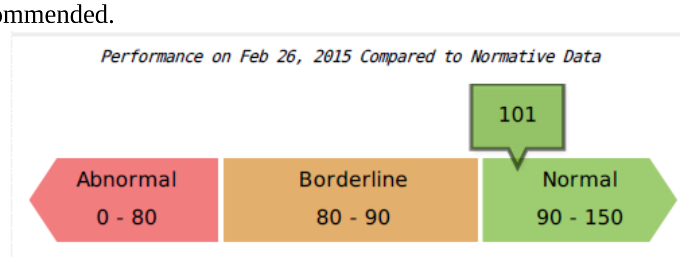



Figure 3: Normative comparison plot displaying performance in the Normal, Borderline or Abnormal ranges.

**3. AFTER-INJURY COMPARISONS TO BASELINE AND NORMATIVE ASSESSMENTS**

The baseline approach to cognitive testing has been demonstrated as a well-established method to evaluating cognitive decline or impairment following a concussion (Louey et al., 2014). An alternative approach to measuring after-injury cognitive performance is to simply compare performance to an age-matched normative sample and determine whether performance is outside of the normal range (termed the normative method). The underlying assumption of the normative method is that elite athletes represent healthy adults who theoretically fit within the healthy distribution of cognitive performance. One shortcoming of this approach is that it is possible, and likely given the large sample of elite athletes playing professional sport, that a proportion of the athletes fall above or below the normal range at baseline. However, if an athletes level of cognitive function is below normal at baseline, after-injury performance will also be below the normal range (and likely lower than their optimal performance) irrespective of whether the athlete is symptomatic as a result of a concussion. As such, the results of the test will not accurately inform a return to play decision. It is also possible that an athlete's baseline cognitive performance is higher than the normative average, and therefore during an after-injury assessment, performance may merely decline into the normal range. Therefore, if a baseline assessment is not available, and a medical officer only has a normative comparison available, it will not be possible to detect that cognition has declined relative to pre-injury levels.

Louey and colleagues (2014) compared the sensitivity and specificity of the baseline and normative data methods on the CBB among a sample of 260 non-injured and 29 recently concussed athletes. Although both methods maintained high correct classification rates and high specificity, the baseline method had higher sensitivity than comparing after-injury performance with normative data. Furthermore, 27.6% of concussed athletes classified as impaired using the baseline method where classified as unimpaired when the normative method was applied.

It is not always possible to obtain a baseline test on each athlete at the beginning of the season. This is particularly the case for amateur athletes where resources are not available to baseline all players. In cases where a baseline assessment is not available, an after-injury assessment can provide an indicator of cognitive performance that would otherwise not be available. Whilst a return to play decision cannot be made on the basis of the assessment, the findings can be used to provide evidence of whether any adverse effects of the head trauma sustained during a match remain present.

## 4. INDIVIDUAL, GROUP AND UNSUPERVISED TESTING

Given technological advances in recent years, it is now possible for a computerised test battery such as the CBB to be administered online, in which case an assessment can be completed in alternative testing environments that are more convenient for the athlete (e.g., at home). This is relevant to baseline testing in particular given the benefit of all players completing a baseline test, and clubs (particularly at non-elite levels) not having resources available to provide a sound testing environment, or adequate supervision, for baseline assessments. Despite the available technology for group-based or unsupervised testing, questions have been raised regarding the validity of assessments not completed in a traditional supervised context, and whether data collected from group-based or unsupervised tests is equivalent to that collected under individual supervision. Cromer and colleagues (2015) completed a two part study where the CBB was administered to undergraduate students across (i) individual supervised versus group supervised conditions, and (ii) individual supervised versus unsupervised conditions. No significant differences were found between supervised assessments completed under individual or group supervision. Furthermore, performance gathered in the unsupervised condition were not significantly different to those the individual supervised condition.

These findings suggest that as technology continues to evolve, consideration could be given to athletes completing baseline assessments in alternate settings that place less burden on club resources. Whilst one of these options in group-based testing where a single supervisor oversees baseline assessments for a group of athletes at a time, another option is unsupervised testing. In the event that an athlete is concussed, a more formal process of supervised testing would be undertaken for the after-injury assessments, with results being interpreted in the context of other formal assessments such as the SCAT3.

Challenges associated with the testing environment also extend to the age of the athletes being tested. For adult athletes, a simple computerised assessment can be completed unsupervised with relative ease, provided the test has inbuilt training modules that provide adequate opportunity for understanding and skill acquisition on the tests being undertaken. Additional validity checks are also needed to evaluate whether the athlete was adhering to the test rules, which enables the medical officer to determine that the completed test was a valid measure of cognition. For younger athletes however, particularly children and adolescents, greater care needs to be taken when considering group-based or unsupervised testing. During the initial stages of testing, particularly during a first attempt at a battery, children may need feedback from a supervisor on the rules of a test. Whilst this tends to be a trained test supervisor in a club environment currently, this may extend to coaches or even parents or caregivers if baseline assessments are completed at home in future. In any case, adequate training must be provided to all supervisors to ensure the level of assistance provided is consistent across athletes.

## 5. OTHER APPLICATIONS OF COGNITIVE TESTING IN SPORT

There are a range of applications of cognitive testing in sport that extend beyond return to injury decisions after concussion. Cognitive tests measure a variety of domains important in elite sport, including but not limited to, processing speed, sustained and divided attention, concentration, working memory, and executive functions such as planning and decision making. These tests are often used to evaluate the success of interventions that target improved cognitive function in typical game scenarios and highly pressurised situations. For example, an AFL player's approach to the final quarter in a game where their team is leading by 45 points may differ to a game where the margin is only 3 points going into the last quarter. Furthermore, the mental state of a player as they walk onto the ground in the lead up to a preliminary or grand final may differ to a regular home-and-away season game. The ability to simulate these types of scenarios for players is paramount, as it enables practice and assimilation to the pressures associated with high-stakes matches. Cognitive tests can be used to assess and provide feedback on cognition and evaluate improvement or deterioration in domains such as attention, concentration and executive functions at critical moments.

## 6. CONCLUSIONS

Whilst a range of assessments are available in assessing concussive symptoms, several assessments are typically used to provide convergent evidence on presenting symptoms and whether a player should return to play. The most commonly used assessment is the SCAT3, and this tool is often accompanied by a computerised cognitive test battery such as the CBB. The CBB has been shown to be a valid and reliable assessment of four core cognitive domains that can be measured in 10 to 15 minutes. The CBB is typically administered in an individual supervised environment, and recent technological advances have enabled this tool to be used in group-based supervised or unsupervised settings. The availability of an online version of the CBB will facilitate ease of assessment at elite levels, but also facilitate collection of baseline assessment data in amateur and community-based sports.

## REFERENCES


Collie, A., Makdissi, M., Maruff, P., Bennell, K., & McCrory, P. R. (2006). Cognition in the days following concussion: comparison of symptomatic versus asymptomatic athletes. *J Neurol Neurosurg Psychiatry*, *77*(2), 241–245. http://doi.org/10.1136/jnnp.2005.073155

Collie, A., Maruff, P., Makdissi, M., McCrory, P. R., McStephen, M., & Darby, D. G. (2003). CogSport: reliability and correlation with conventional cognitive tests used in postconcussion medical evaluations. *Clinical Journal of Sport Medicine : Official Journal of the Canadian Academy of Sport Medicine*, *13*(1), 28–32. http://doi.org/10.1097/00042752-200301000-00006

Concussion in Sport Group (2013). Sport Concussion Assessment Tool - 3rd Edition. *British Journal of Sports Medicine*, *47*, 259.

Covassin, T., Elbin, R. J., Stiller-Ostroski, J. L., & Kontos, A. P. (2009). Immediate post-concussion assessment and cognitive testing (ImPACT) practices of sports medicine professsionals. *Journal of Athletic Training*, *44*, 639-644.

Cromer, J. A., Schembri, A. J., Harel, B. T., & Maruff, P. (2015). The nature and rate of cognitive maturation from late childhood to adulthood. *Frontiers in Psychology*, *8*, 1-12.

Cromer, J. A., Harel, B. T., Yu, K., Valadka, J. S., Brunwin, J. W., Crawford, C. D., … Maruff, P. (2015). Comparison of Cognitive Performance on the Cogstate Brief Battery When Taken In-Clinic, In-Group, and Unsupervised. *The Clinical Neuropsychologist*, *29*(4), 542–558. http://doi.org/10.1080/13854046.2015.1054437

Eckner, J. T., Kutcher, J. S., & Richardson, J. K. (2010). Pilot evaluation of a novel clinical test of reaction time in National Collegiate Athletic Association Division i football players. *Journal of Athletic Training*, *45*(4), 327–332. http://doi.org/10.4085/1062-6050-45.4.327

Gleeson, M. (2016). Brisbane lions defender Justin Clarke retires at 22 from concussion. http://www.theage.com.au/afl/afl-news/afl-2016-brisbane-lions-defender-justin-clarke-retires-at-22-from-concussion-20160330-gnuju1.html

Louey, A. G., Cromer, J. A., Schembri, A. J., Darby, D. G., Maruff, P., Makdissi, M., & McCrory, P. R. (2014). Detecting Cognitive Impairment After Concussion: Sensitivity of Change From Baseline and Normative Data Methods Using the CogSport/Axon Cognitive Test Battery. *Archives of Clinical Neuropsychology*, *29*(5), 432–441. http://doi.org/10.1093/arclin/acu020

Makdissi, M., Collie, A., Maruff, P., Darby, D. G., Bush, A. I., McCrory, P. R., & Bennell, K. (2001). Computerised cognitive assessment of concussed Australian Rules footballers. *British Journal of Sports Medicine*, *35*(5), 354–60. http://doi.org/10.1136/bjsm.35.5.354

Makdissi, M., Darby, D. G., Maruff, P., Ugoni, A., Brukner, P., & McCrory, P. R. (2010). Natural history of concussion in sport: markers of severity and implications for management. *Am J Sports Med*, *38*(3), 464–471. http://doi.org/0363546509349491 [pii]10.1177/0363546509349491

# LAW AND ORDER FOR SPORTS LEAGUES: SIMULATING DRAFT PENALTIES AND LEAGUE DYNAMICS USING SPORTS SYNTHESIS

Athol Whitten[a,d], Geoff Tuck[b], and Robert Macdonald[c]
[a] *Mezo Research, Australia*
[b] *CSIRO Data 61 and CSIRO Oceans and Atmosphere*
[c] *University of Melbourne, Australia*
[d] *Corresponding author: athol.whitten@mezo.com.au*

## Abstract

Serious breaches of league regulations by sporting clubs lead to management interventions such as monetary fines and draft penalties. Draft penalties remove some number of a club's promised draft selections over one or more seasons. Recent prominent examples include the draft penalties imposed upon Australian Football League (AFL) clubs Carlton, Adelaide, and Essendon in 2002; 2013; and 2014 respectively. Annual draft systems are the principal method by which clubs in closed sporting leagues recruit amateur players. Reverse-order drafts, such as those used by the National Football League (NFL), National Basketball Association (NBA), Major League Baseball (MLB) and the AFL, allow clubs with the poorest win-loss record in a season to access the most highly-rated amateur players. Penalising a club through the player draft implicitly assumes that high draft picks are valuable and will ultimately improve club performance, and that removal of draft picks will reduce the probability of club success in the following years. Currently, there are (i) no mechanisms to determine the degree to which these draft penalties might reduce a club's probability of success, and (ii) no clear articulation from management bodies of the desired magnitude of the reduction in the probability of success. Here, we use a dynamic simulation model 'Sports Synthesis' to capture the key components of a win-maximising sport league, such as the AFL. Sports Synthesis incorporates an amateur player draft (reverse-order and others), player productivity, between-team competition and draft choice error (i.e. the ability of clubs to determine player quality in the draft). We show how the magnitude of the penalty (in terms of the reduction in the probability of success) is dependent on the number and position of draft selections removed, key league characteristics (such as draft choice error) and the ladder position (or team productivity) of the club when penalties are applied.

*Keywords*: Simulation modelling, draft penalties, league management, management strategy evaluation

# TRAINING BADMINTON PLAYERS AND KNOWING WHEN TO THROW IN THE TAU

Minh Huynh [a,b], Kelly Marshall [a]
[a] *Swinburne University of Technology*
[b] *Corresponding author: mhuynh@swin.edu.au*

## Abstract

The aim of the present study was to examine general tau theory in relation to the sport badminton. We cover the three main concepts of motion-gaps, tau-coupling and time-to-contact. The second half of this paper examines how tau theory can be utilised by coaches and trainers in training their athletes. Specifically, we discuss anticipatory training, time-to-contact theory, skills acquisition training for badminton program, and the required velocity model. Finally, we suggest scenarios and experiments that can be researched in future studies to discover an ideal training method that combines both physical training and cognitive recognition.

**Keywords**: Tau theory, time-to-contact, badminton, skill acquisition

## 1. INTRODUCTION

Knowing how to anticipate approaching objects and when they will reach a desired goal or target is a crucial survival skill for any living organism to possess (Hancock and Manser, 2009). In particular, mastering anticipatory skills aide in competitive fast paced sports, where movement has to be controlled by perceiving what is likely to occur next (Kayed and Van der Meer, 2009). For example, catching a cricket ball or striking a badminton shuttle requires precise prospective control of the interceptive action and must be prepared in advance to allow the body time to organize. It can therefore be suggested that prospective control is dependent on perceptual information guiding the action so that an extrapolation of movements can be made into the future (Kayed and Van der Meer, 2009; Lee, 2005; Von Hofsten, 2007). This theory of prospective control is commonly known as general tau theory and has been associated with various forms of development and anticipation. This paper will examine tau in detail and explain how it can be used by badminton players for skill acquisition training. In addition, this paper will also summarise motion-gaps, tau-coupling and time-to-contact concepts.

## MOTION GAPS AND BADMINTON

The concept of a motion gap can generally be described as changing the gap between a current state and a goal state in a given event (Lee, 2005), and is typically linked by an effector. If we take a badminton rally as an example, the current state could be the position of the player's racket, the effector would be the racket itself, and the goal state would be the shuttlecock. This is an example of closing a *distance* motion gap during a game of badminton. Other examples include *force* motion gaps (e.g. a player's current force and the force required to execute a satisfactory jump) and *angular* motion gaps (e.g. the distance between a player's current field of vision and the direction of the shuttlecock).

In any given game of badminton, or other racket sports, it is unlikely that an action will only have one case of closing a single motion gap. In reality, we need to control several motion-gaps at the same time to perform a successful action. For example, executing a jump smash in a game of badminton is an example of closing several motion-gaps at the same time. The player needs to coordinate the closure of vision-shuttlecock, racket-shuttlecock, and jump-court motion-gaps to perform the jump smash adequately. Performing a jump smash, and at precisely the right time, can be a difficult task, especially for such a fast paced sport. To explain how performance with such a strict temporal constraint can be executed based exclusively on raw visual information, it would be ideal to examine the concept of tau and motion-gaps.

## TAU-COUPLING AND BADMINTON

Lee (1998) suggests that tau, as a single type of temporal variable of a changing motion-gap, would provide sufficient information for controlling the closure of said motion-gap. The notion of tau is based on Gibson's (1966) work on ecological invariants in visual flow fields. Tau of a motion-gap is the time-to-closure of the motion-gap at its current closure-rate:

$$\boldsymbol{\tau}(\mathrm{x},\mathrm{t}) = \frac{\mathrm{x(t)}}{\dot{\mathrm{x}}\mathrm{(t)}} \qquad (1)$$

where τ(x, t) represents the tau of the motion-gap at time t. Because tau is a measure on any motion-gap of any dimension, it explains how a single type of temporal variable can account for controlling the closure of perceptual information from different dimensions of motion-gaps.
The notion of tau-coupling refers to when two taus are coupled over a period of time if they remain in constant proportion during that time (Lee, 2005):

$$\boldsymbol{\tau}(x,t) = K\boldsymbol{\tau}(y,t) \qquad (2)$$

where τ(x)and τ(y)represent the taus of two gaps, K represents the coupling constant, and t represents time. For example, consider a badminton player intercepting a shuttlecock at point X (see Figure 1).

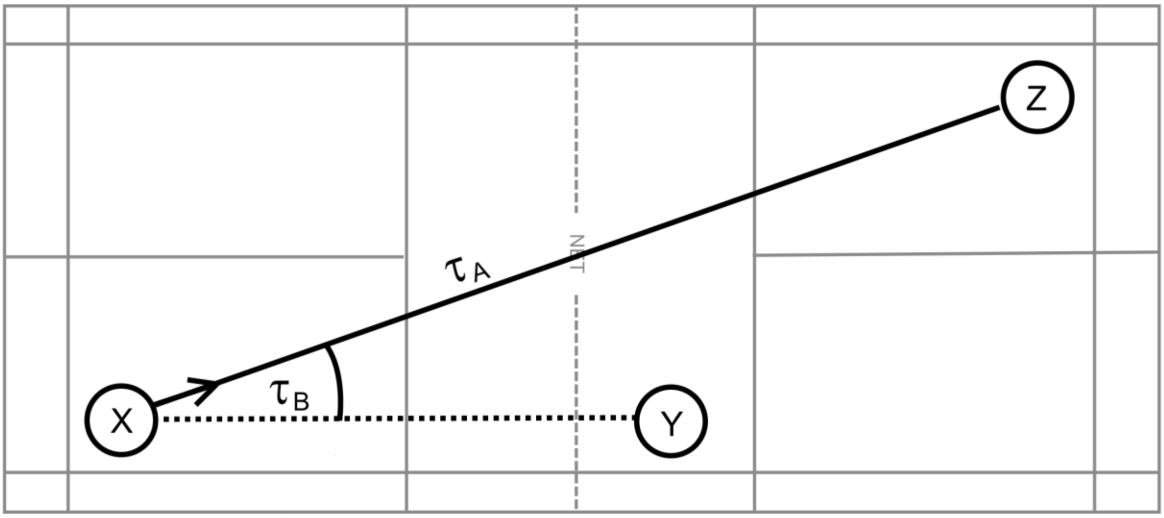



Figure 1: Tau coupling example for scenario 1

The player notices their opponent at point Y so decides to hit the shuttlecock to a location that would be the most difficult for the opponent to close (i.e. point Z), given their current location. To execute this action successfully, the badminton player needs to control the closure of the shuttle trajectory motion-gap, A, as well as the angular motion-gap, B, between the shuttlecock pathway and his opponent at point Z simultaneously. Utilising the equation above (2), this scenario can be expressed symbolically as:

$$\boldsymbol{\tau}_B = K\boldsymbol{\tau}_A \qquad (3)$$

Also consider a second scenario where a badminton player running to the left side of the court to return his opponent’s stroke from point A (refer to figure 2). To execute this properly, the player has to control the closure of both gaps from the player to the shuttlecock, X, and the shuttlecock trajectory route, TR, simultaneously. This can be expressed symbolically as:

$$\boldsymbol{\tau}_X = K\boldsymbol{\tau}_{XTR} \qquad (4)$$

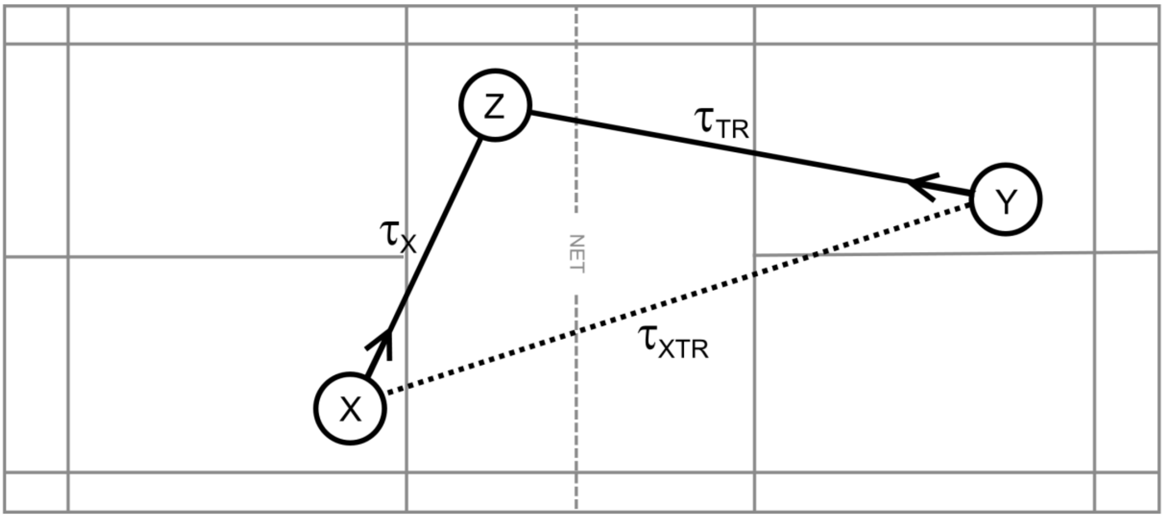


Figure 2: Tau coupling example for scenario 2

**TIME-TO-CONTACT IN BADMINTON**

Another important concept in understanding tau and tau-coupling, is time-to-contact (Tc). The Tc theory suggests that living organisms are able to perceive using visual information to judge distance and speed with respect to time (Bootsma and van Wieringen, 1990). In determining Tc, we can utilise tau (mentioned above) as our optical invariant because it accounts for perceptual information coming from different dimensions (in this case: speed and distance).

Tau relates the optical size of an object to its rate of expansion and can be expressed as:

$$\tau = \theta/\delta\theta/\delta t \qquad (5)$$

where θ represents the angle of extension of the object in radians and $\delta\theta/\delta t$ represents the rate of expansion. Under Tc conditions, tau specifies that an object must move with constant velocity. As an example, consider a game of badminton played by experts. Instinctively, players are able to return shots of up to 421 km/h based on collecting visual information from their opponent's actions. This plays an important role in the development of skill acquisition and training and will be discussed later.

In summary, we can infer a number of conclusions regarding the nature in which people and animals use tau to guide their movements:

1. Rather than breaking down actions into discrete movements (ie muscle movements), it is useful to consider actions in terms of controlling the closure of motion-gaps
2. Actions are initiated when a certain threshold of tau (time to closure of the motion gap) is reached
3. People and animals continuously assess the time to closure of motion gaps (tau). If tau is inadequate to complete the desired task (ie cross the court to be in the right position to return a serve), the rate of closure can be changed, ie "run faster", "turn more sharply", "apply more force with the racket", etc.
4. Complex actions involve multiple motion-gaps, which can be controlled by tau-coupling, or linking two or more motion-gaps so that they are closed simultaneously.

By combining visual information with an internal forward model of the external state, the relationship between position and Tc becomes easily predictable by most animals (de Azevedo Neto and Teixeira, 2009). Even in extremely fast-paced sports, humans are able to estimate time windows for successful performances in the order of a few milliseconds (Regan, 1992; Tresilian, 1993; Land and Mcleod, 2000; de Azevedo Neto and Teixeira, 2009). Therefore, it would be possible to improve a badminton player's decision making capabilities utilising a visual based training model based on the concepts of tau and time-to-contact.

## 2. TAU AND SKILL DEVELOPMENT

Traditional skills training programs for badminton players focus primarily on the development of motor skills, such as the mechanics of each type of shot. Cognitive and decision-making skills are generally left to develop naturally during games, despite the suggestion that these skills are almost as important as the motor skills in badminton (Blomqvist, Luhtanen & Laakso, 2001). Therefore it is beneficial to incorporate aspects of cognitive and decision-making skills alongside traditional training approaches.

In this next section we will discuss how general tau theory can be used to enhance training programs, improve a player's skills, techniques and overall in-game performance.

### USING TAU TO ANTICIPATE TIME

Players can estimate when a shuttlecock will land at a location on the court by making use of tau and the knowledge of Tc. However, this does not mean that the shuttlecock will definitely come into contact with the racket, for this requires an action on the part of the player that must be geared towards tau (Bootsma and van Wieringen, 1990). The synchronisation between movement and visual information can generally be achieved be executing the same movement repeatedly, with as little variation as possible. A classic example of this is Bootsma and van Wieringen's (1990) study relating to timing forehand drives in table tennis. The researchers suggest that with enough repetition, players have automatically geared themselves towards making the optimal forehand drive. In this sense, the only challenge players' face is deciding *when* to initiate the drive. The same principle can be applied to our badminton example. Players could wait until tau reaches a critical value which would indicate the optimal time to execute the movement. Using this technique coaches can train their athletes to identify the ideal time to execute an action for any situation during a badminton game. For example, coaches may hit a large number of shuttlecocks to the same location on the court with as little variability as possible. Players would then keep returning these shuttles until they discover the optimal time to initiate the movement execution process (assuming they have already decided the type of shot they will make). This approach allows players to develop visual cues which act as signals during real game situations.

### TAU-COUPLING TO REFINE COMPLEX MOVEMENTS

The ability to execute high end techniques in badminton require controlling the closure of a set of motion-gaps in a specific manner (refer to the example regarding the execution of the jump smash). Lee (2005) suggests that skilled movements may be acquired by coupling the taus of the motion-gaps onto the taus of other motion-gaps and further onto tau-G guides (refer to Lee, 2005 for a detailed explanation of tau-G). As an example, badminton players learn to adjust their movement around the court by absorbing the visual information provided by their opponents. When following the shuttlecock with their eyes, players learn to not only sense the motion-gap between their gaze and the shuttle, but also learn how to control the optimal amount of time spent (in the order of a few milliseconds) in gazing at the shuttle and the decision that would follow. Of course, no real life event is ever repeated, in the same sense that no badminton rally is ever identical. Thus, players learn to improve their skills through constant adjustments of the calibration process of regulating power to the muscles on the basis of prospective sensory information about the taus of motion-gaps (Lee, 2005).

### THE REQUIRED VELOCITY MODEL

Peper et al. (1994) suggest a required velocity model which examines Tc and how it can be used to intercept moving objects. The model itself specifies how an individual utilises visual information regarding his environment continuously to control the hand's acceleration and match the required velocity needed to intercept an object. Generally, the current hand velocity at a given instant *t* can be increased or decreased for the hand to move at the required velocity (Davids, Button and Bennett, 2008) to intercept an object. In a study conducted by Peper et al. (1994) the researchers examined the required velocity to catch a ball approaching a person at a specific speed. The equation they suggested for the required velocity model can be derived as:

$$\ddot{X}_h = \alpha \dot{X}_{h\,req} - \beta \dot{X}_h \qquad (6)$$

$$\text{with } \dot{X}_{h\,req} = \frac{X_h - X_b}{Tc_1} \qquad (7)$$

where $\ddot{X}_h$, $\dot{X}_{h\,req}$, and $\dot{X}_h$, are the hand's current acceleration, current velocity, and required velocity respectively, and α and β are constants, and where $X_h$, $X_b$, and $Tc_1$ are the hand's current position, the projection of the ball's current position on the hand-movement axis, and the first-order Tc between the ball and the hand-movement axis. Using this model, we can derive a similar equation for the required velocity a badminton player needs to intercept a shuttlecock hit in-game:

$$\text{Required velocity} = \frac{X_{hand} - X_{shuttle}}{Tc_1} \qquad (8)$$

Because badminton is such a high velocity game, players constantly have to make very quick in-game decisions. Macquet and Fleurance's (2007) analysis of decision-making during badminton matches suggests that there are times when the player must give up a shot in order to win the overall set. Therefore a key decision-making skill is whether to attempt the return, or give up the shot. Knowing which shots are possible from their current position (specifically hand position) and the required velocity they would need to reach that point could potentially save them a lot of energy from unnecessary movement and provide a tactical advantage. Of course, players wouldn't have the time to be calculating required velocity in-game, however this model may be effective during the training period of the athlete. Coaches and trainers may train players to mentally recognise shuttle velocity in relation to their current position, such that in a real game situation, they would recognise which shots they should disregard to conserve energy for the next point.

## 3. CONCLUSION

The first section of this paper attempted to summarise three main concepts of tau theory (motion-gaps, tau-coupling and time-to-contact) in relation to the sport badminton. The second section examined various ways of training badminton players while utilising a general tau theory as a background. Overall, the theory relates to the movement guidance of animals and how they acquire understanding of their surrounding environments through perceiving visual information. Naturally, anticipating when an object approaches another object is in inherent trait most organisms possess. However, it would be interesting to observe if coaches could utilise general tau and time to contact theory when training the anticipatory skills of athletes. Perhaps if we have more coaches throwing in tau theory during their training sessions, then we may have fewer players throwing in the actual towel during their games.

**Acknowledgements**
This paper is based on work from the primary author's Master's dissertation. A special thanks to the author's principal supervisor, Associate Professor Anthony Bedford, and his assistance with the research.

**References**

Blomqvist, M., Luhtanen, P, & Laakso, L. (2001). Comparison of two types of instructions in badminton. *European Journal of Physical Education, 6,* 139-155.

Bootsma, R. J., & van Wieringen, P. C. W. (1990). Timing an attacking forehand drive in table tennis. *Journal of experimental psychology: Human perception and performance, 16*, 21-29.

Davids, K., Button, C., & Bennett, S. (2008). *Dynamics of skill acquisition: A constraints-led approach*. South Australia: Human Kinetics.

de Azevedo Neto, R. M., & Teixeira, L. A. (2009). Control of interceptive actions is based on expectancy of time to target arrival. *Exp Brain Res*, 199, 135-143.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

Hancock, P. A., & Manser, M. P. (2009). Time-to-contact: More than Tau alone. *Ecological Psychology*, 9, 265-297.

Kayed, N. S., & Van der Meer, A. L. H. (2009). A longitudinal study of prospective control in catching by full-term and preterm infants. *Exp Brain Res, 194,* 245-258.

Land, M. F., & McLeod, P. (2000). From eye movements to actions: How batsmen hit the ball. *Nat Neurosci, 3,* 1340-1345.

Lee, D. N. (1998). Guiding movement by coupling taus. *Ecological Psychology, 10,* 221-250.

Lee, D. N. (2005). Tau in action in development. In: Reiser, J. J., Lockman, J. J., & Nelson, C. A. (2005). *Action as an organiser of learning and development*. Lawrence Erlbaum Associates, Mahwah, pp 3-47.
Macquet, A.C., & Fleurance, P. (2007). Naturalistic decision-making in expert badminton players. *Ergonomics, 50*, 1433-1450.
Peper, C. E., Bootsma, R. J., Mestre, D. R., & Bakker, F. C. (1994), Catching balls: How to get the hand to the right place at the right time. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 591-612.
Proffitt, D. R., & Caudek, C. (2003). Depth perception and the perception of events. In: Weiner, I. B., Healy, A. F., Freedheim, D. K., Proctor, R. W., & Schinka, J. A. (2003). *Handbook of psychology: Experimental psychology*. John Wiley & Sons, Inc, New Jersey, pp 213-236.
Regan, D. (1992). Visual judgments and misjudgements in cricket, and the art of flight. *Perception, 21*, 91-115.
Tresilian, J. R. (1993). Four questions of time to contact: A critical examination of research on interceptive timing. *Perception, 22*, 653-680.
Von Hofsten, C. (2007). Action in development. *Dev Sci, 10*, 55-60.

# THE CONVEX HULL OF A BALLPLAYER

Gerry Myerson[a,b]
*a Macquarie University*
*b Corresponding author: gerry.myerson@mq.edu.au*

## Abstract

Athletes have their ups and downs. We form expectations of their level of performance based on their best years, and may consider they have not lived up to their potential in their other years. We propose a way to quantify an athlete's potential, and the extent to which the athlete has fallen short of this potential, based on the familiar concept of the convex hull of a finite point-set. We present our method in the context of yearly home run production in American major league baseball.

**Keywords**: Potential, convex hull

## 1. THE MODEL

Babe Ruth hit 59 home runs during the 1921 (American major league) baseball season, a record at the time. In 1927, he hit 60. One might say that he established that, during the period 1921-27, he had the potential to hit 59 or 60 home runs every season. He didn't actually do this; in fact, his home run totals in the seasons 1922-26 were 35, 41, 46, 25, and 47. In this paper, we present a candidate definition for a player's home run potential at any point in his career, and a corresponding measure of the extent to which a player falls short of his potential; we apply our definition and measure to the members of baseball's "400 Home Run Club", and to several other players who had notable successes in hitting home runs; we analyze the careers of some of the outliers in the study, the players who came the closest to, or fell the farthest short of, fulfilling their home run potential.

Before we get stuck into it, we wish to stress two things. One, although we will only look at baseball, and only at home run hitting, we hope it will be clear that the mathematical tools we use will be widely applicable to other sports and indeed to other human endeavors. In essence, they will apply whenever there is something that can be counted, with the counts fluctuating from one time period to the next. Second, although we speak of players failing to meet their potential, we mean no criticism of the players. There are many reasons for failing to meet potential that have nothing to do with personal failings of the ballplayer; playing time missed to injuries, or to military service, and just plain luck, to mention a few.

To create our definition of a player's home run potential for each season in his career, we plot the points $(i, y_i)$, where $y_i$ is the number of home runs hit by the player in the $i$th year of his career. A career is taken to include every year from the first year the individual plays in the major leagues to the last, even if that includes one or more years during which he does not actually play. We also plot the point (0, 0), and, if the number of years in the career is $r$, the point $(r+1, 0)$. We call this the player's *Home Run Plot*. For example, Ralph Kiner played in the major leagues every season from 1946 to 1955, with yearly home run totals

23, 51, 40, 54, 47, 42, 37, 35, 22, 18

so his Home Run Plot consists of the points,

(0, 0), (1, 23), (2, 51), (3, 40), (4, 54), (5, 47), (6, 42), (7, 37), (8, 35), (9, 22), (10, 18), (11, 0).

We note here that all the seasonal home run figures used in this paper are taken from the Baseball-Reference website. We then compute the convex hull of the plotted points (or we ask a computer algebra package to do this computation for us). In our example, this is the convex polygon with vertices at

(0, 0), (2, 51), (4, 54), (8, 35), (10, 18), (11, 0)

The main properties of the convex hull are that it is convex (so it contains the line segment joining any two of its points), its vertices are a subset of the plotted points, and it contains all of the plotted points. It clearly has a lower boundary, the line segment joining (0, 0) to ($r$+1, 0), and an upper boundary, consisting of all the line segments forming the remainder of its perimeter (we are not interested in the degenerate case of the ballplayer who hits no home runs during his career).

We now define a player's home run potential in year $i$ of his career to be the $y$-coordinate of the point on the upper boundary of the convex hull with $x$-coordinate $i$. Thus, Kiner's home run potential in the seventh year of his career would be the number $y$ such that (7, $y$) is on the line segment joining (4, 54) and (8, 35), which is 39.75.

One might object that no one has ever hit three-fourths of a home run. The objection (mostly) goes away when we sum the home run potential over all the years of a player's career, to get the player's career potential home runs; for brevity's sake, we refer to this sum as the player's *Hull*. The sum will be an integer or, at worst, a half-integer:

Theorem. A player's Hull is equal to the area of the convex hull of his Home Run Plot, and can be computed by the formula,

One-half the sum of $a_i b_{i+1} - a_{i+1} b_i$, $i$ running from 0 to $s - 1$.

where $(a_i, b_i)$, $i = 0, \ldots, s$, are the vertices of the convex hull, with $0 = a_0 < a_1 < \ldots < a_s$.

The equivalence of the area of the convex hull and the sum of the ordinates of points on the hull's upper boundary corresponding to integer abscissas can be seen by partitioning the hull into $r - 1$ trapezoids and two triangles by drawing vertical line segments joining each point $(i, 0)$, $i = 1, \ldots, r$, to the upper boundary and then applying formulas for the areas of trapezoids and triangles. The displayed formula is a well-known formula for the area of a polygon in the plane, in terms of the coordinates of its corners.

For Kiner, we find that his Hull is 396, as compared with his actual career total of 369 home runs. As we shall see, Kiner has a remarkably small gap between his Hull and what we may call his Actual.

## 2. THE 400 HOME RUN CLUB

Baseball's "400 Home Run Club" consists of those ballplayers whose career home run totals come to 400 or more. Through the 2015 season, the Club had 52 members. Four of these were active during the 2015 season, and we have omitted them from this study. For the remaining 48 ballplayers, we tabulate their career home run totals ("Actual"), their career potentials as given by the area of the convex hull ("Hull"), the difference between these two numbers as a measure of the player's shortfall ("Diff"), and their actual as a percentage of their potential ("Percentage", calculated as 100 times Actual divided by Hull), in Table 1.

Most of the percentages follow a fairly flat distribution from a low of 69.69 (Andre Dawson) to a high of 86.92 (Adam Dunn). Their are two outliers at the bottom of the range, and they deserve closer attention: Ted Williams, 61.84, and Darrell Evans, 58.02.

Here are Ted Williams' year-by-year home run counts:

31, 23, 37, 36, 0, 0, 0, 38, 32, 25, 43, 28, 30, 1, 13, 29, 28, 24, 38, 26, 10, 29

Here and below we underline the numbers that correspond to vertices on the convex hull; keep in mind that the convex hull also has vertices at (0, 0) and (for Williams) at (23, 0). The string of zeros in years five through seven reflect Williams' military service during the Second World War. The model suggests that Williams could have hit 118 home runs during those three seasons. Williams also served during the Korean War, playing only six games in 1952, his 14th season, and 37 games in 1953 (out of a scheduled 154 games each season). Also contributing to Williams' low Percentage is his outstanding 19th season, 1957, when he hit 38 home runs. Only Bonds, Aaron, and Ruth have hit so many home runs in a season that late in their career.

Here are Darrell Evans' home run counts:

0, 0, 12, 19, 41, 25, 22, 11, 17, 20, 17, 20, 12, 16, 30, 16, 40, 29, 34, 22, 11.

Evans only played 12 games (out of a scheduled 162) in each of his first two seasons, which explains the zeros in those first two years. Nothing explains the two years of 40+ home runs, separated by 11 seasons during which Evans averaged only 19 home runs. Evans is a true outlier.

Adam Dunn's career, at the top of the Percentage list at 86.92, is also worth a look:

19, 26, 27, 46, 40, 40, 40, 40, 38, 38, 11, 41, 34, 22.

Only twice in his 14-year career did Dunn fall more than five home runs short of his potential. His poor showing in his 11th season, 2011, may be explained by his return to the field too soon after an early season appendectomy.

## 3. OUTSIDE THE CLUB

We have made an unsystematic selection of 14 ballplayers who did not hit 400 or more home runs, but who did have one or more seasons with a large number of home runs. Since this was not an exhaustive study, the only thing to be safely concluded from the data in Table 2 is that these players show more variability than the group of 48. Several of these players sport percentages lower than those of Ted Williams and Darrell Evans, with the most extreme case being that of Hank Greenberg (46.23):

0, 0, 0, 12, 26, 36, 1, 40, 58, 33, 41, 2, 0, 0, 0, 13, 44, 25.

Greenberg played one game in the major leagues in 1930, spending the rest of that season and all of the next two in the minor leagues. This explains the zeros at the beginning of his career. In his seventh year, he only played 12 of a scheduled 154 games before a season-ending injury left him with only one home run to show for 1936. He was drafted into the Army in May, 1941, just 19 games and two home runs into the season. He was discharged from the Army on 5 December 1941, and re-enlisted immediately after the US entered the war two days later. He was not to play baseball again until he was discharged and rejoined his team halfway through the 1945 season. These circumstances, combined with the extraordinary 58 home runs in his ninth year, push him to the top of the "Diff" list, with 385 career home runs fewer than the convex hull model. Ted Williams is a distant second, at 321.5.

We note in passing that Joe DiMaggio and Johnny Mize each lost three years to military service in mid-career, contributing to their low percentages.

At the other end of the Percentage list, with 93.18, is Ralph Kiner, whose year-by-year totals we have already seen. Curiously, Kiner's career, like Greenberg's, was severely affected by military service and medical problems. The difference is that Kiner served in the military before his major league career began. He was already 23 when he came to the majors, and didn't have the slow start to his career that some others did. Instead of coming back from a serious injury, as Greenberg did, Kiner was driven from the game at age 32 by a bad back. These factors account, in part, for his coming so close to achieving his full potential as a home run hitter.

## 4. WAY OUTSIDE THE CLUB

We conclude with the curious cases of Johnny Cooney and Nick Altrock. Cooney was a major league pitcher for ten years. He then went down to the minors for four years, and returned to the majors as an outfielder for another ten-year stint. During these 24 years, he hit two major league home runs, both in the same season. This gives him a Hull of 25, and a Percentage of 8, most likely the second lowest figure in baseball history.

The lowest figure probably belongs to Nick Altrock. He pitched in 1898, and again from 1902 to 1909. After that, he would occasionally be used in one or more games at the end of the season, mostly as a stunt. His last appearance was in 1933, when he was 56 years old. During this span of 36 years, he hit two home runs, one in 1904 and the other in 1918. His Hull is 25.5, his Percentage, 7.84.

| | | Actual | | Hull | | Diff | | Pct |
|---|---|---|---|---|---|---|---|---|
| Barry Bonds | | 762 | | 1034.5 | | 272.5 | | 73.66 |
| Hank Aaron | | 755 | | 902 | | 147 | | 83.7 |
| Babe Ruth | | 714 | | 946.5 | | 232.5 | | 75.44 |
| Willie Mays | | 660 | | 923 | | 263 | | 71.51 |
| Ken Griffey, Jr | | 630 | | 843 | | 213 | | 74.73 |
| Jim Thome | | 612 | | 744.5 | | 132.5 | | 82.2 |
| Sammy Sosa | | 609 | | 799.5 | | 190.5 | | 76.17 |
| Frank Robinson | | 586 | | 790.5 | | 204.5 | | 74.13 |
| Mark McGwire | | 583 | | 879.5 | | 296.5 | | 66.29 |
| Harmon Killebrew | | 573 | | 799 | | 226 | | 71.71 |
| Rafael Palmeiro | | 569 | | 663.5 | | 94.5 | | 85.76 |
| Reggie Jackson | | 563 | | 783 | | 220 | | 71.9 |
| Manny Ramirez | | 555 | | 672.5 | | 117.5 | | 82.53 |
| Mike Schmidt | | 548 | | 649.5 | | 101.5 | | 84.37 |
| Mickey Mantle | | 536 | | 700.5 | | 164.5 | | 76.52 |
| Jimmie Foxx | | 534 | | 759 | | 225 | | 70.36 |
| Willie McCovey | | 521 | | 737.5 | | 216.5 | | 70.64 |
| Frank Thomas | | 521 | | 706.5 | | 185.5 | | 73.74 |
| Ted Williams | | 521 | | 842.5 | | 321.5 | | 61.84 |
| Ernie Banks | | 512 | | 661.5 | | 149.5 | | 77.4 |
| Eddie Mathews | | 512 | | 625.5 | | 113.5 | | 81.85 |
| Mel Ott | | 511 | | 678 | | 167 | | 75.37 |
| Gary Sheffield | | 509 | | 699 | | 190 | | 72.82 |
| Eddie Murray | | 504 | | 595 | | 91 | | 84.71 |
| Lou Gehrig | | 493 | | 658.5 | | 165.5 | | 74.87 |
| Fred McGriff | | 493 | | 579.5 | | 86.5 | | 85.07 |
| Stan Musial | | 475 | | 616.5 | | 141.5 | | 77.05 |
| Willie Stargell | | 475 | | 672 | | 197 | | 70.68 |
| Carlos Delgado | | 473 | | 565 | | 92 | | 83.72 |
| Chipper Jones | | 468 | | 586.5 | | 118.5 | | 79.8 |
| Dave Winfield | | 465 | | 636 | | 171 | | 73.11 |
| Jose Canseco | | 462 | | 653 | | 191 | | 70.75 |
| Adam Dunn | | 462 | | 531.5 | | 69.5 | | 86.92 |
| Carl Yastrzemski | | 452 | | 670 | | 218 | | 67.46 |
| Jeff Bagwell | | 449 | | 535.5 | | 86.5 | | 83.85 |
| Vladimir Guerrero | | 449 | | 539 | | 90 | | 83.3 |
| Dave Kingman | | 442 | | 606.5 | | 164.5 | | 72.88 |
| Jason Giambi | | 440 | | 570 | | 130 | | 77.19 |
| Paul Konerko | | 439 | | 545.5 | | 106.5 | | 80.48 |
| Andre Dawson | | 438 | | 628.5 | | 190.5 | | 69.69 |
| Juan Gonzalez | | 434 | | 577 | | 143 | | 75.22 |
| Andrew Jones | | 434 | | 568 | | 134 | | 76.41 |
| Cal Ripken | | 431 | | 564 | | 133 | | 76.42 |
| Mike Piazza | | 427 | | 508 | | 81 | | 84.06 |
| Billy Williams | | 426 | | 541.5 | | 115.5 | | 78.67 |
| Darrell Evans | | 414 | | 713.5 | | 299.5 | | 58.02 |
| Alfonso Soriano | | 412 | | 562 | | 150 | | 73.31 |
| Duke Snider | | 407 | | 516.5 | | 109.5 | | 78.8 |

Table 1

| Player | Actual | Hull | Diff | Pct |
|---|---|---|---|---|
| Frank Howard | 382 | 501 | 119 | 76.25 |
| Ralph Kiner | 369 | 396 | 27 | 93.18 |
| Joe DiMaggio | 361 | 622.5 | 261.5 | 57.99 |
| Johnny Mize | 359 | 641 | 282 | 56.01 |
| Luis Gonzalez | 354 | 594.5 | 240.5 | 59.55 |
| Hank Greenberg | 331 | 716 | 385 | 46.23 |
| Derrek Lee | 331 | 472.5 | 141.5 | 70.05 |
| Shawn Green | 328 | 420 | 92 | 78.1 |
| Ted Kluszewski | 279 | 439.5 | 160.5 | 63.48 |
| Roger Maris | 275 | 405.5 | 130.5 | 67.82 |
| Hack Wilson | 244 | 364 | 120 | 67.03 |
| Brady Anderson | 210 | 400 | 190 | 52.5 |
| Jim Gentile | 179 | 268 | 89 | 66.79 |
| Richard Hidalgo | 171 | 254.5 | 83.5 | 67.19 |

Table 2

**References**

Baseball-Reference.com, http://www.baseball-reference.com

# RATING THE ATTACKING PERFORMANCE OF A NON-WICKET TAKING BOWLER IN LIMITED OVERS CRICKET

Paul J. Bracewell [a,b,d], Martin Coomes [c], James D. Nash [a,b], Denny H. Meyer [c], Samuel J. Rooney [a]

[a] *DOT loves data*
[b] *Victoria University of Wellington*
[c] *Swinburne University of Technology*
[d] *Corresponding author:* paul@dotlovesdata.com

## Abstract

A common measure used in cricket to summarise the effectiveness of a bowler's prowess is the bowling average. This metric is calculated by dividing the number of runs conceded by the bowler by the number of batsmen that they individually dismissed. However, when the bowler does not take a wicket this metric is undefined within that innings, which can lead to incorrect interpretations of a bowler's contribution within the team context. As cricket is a team game, bowlers work in pairs to create pressure to take wickets for the team. Therefore, metrics that reflect this team contribution are more informative and reflective of future performance.

Reject inference is used to estimate the likelihood of a bowler taking a wicket within an innings based on ball by ball data. This approach is commonly used within the financial services sector to build credit risk scorecards due to the non-random absence of behavioural attributes, specifically from applicants with poor credit history. An extensive review of approaches indicates that Memory-based Reasoning delivers superior correlation with future performance and a model structure of practical significance.

Gradient Boosting was used to select scorecard variables summarising individual bowling performance variables that were both practically and statistically significant. The model indicates that within the limited overs game, bowlers who deliver a higher number of dot balls and have a lower economy rate are most likely to take a wicket. The inferred metric for measuring the attacking performance is validated by comparing observed and future performance. As an attacking metric can be defined for every bowler per innings, this enables holistic bowling performances to be monitored for changes in form and fitness.

*Keywords*: Memory-based Reasoning, Reject Inference, Parcelling, Gradient Boosting

## 1. INTRODUCTION

Cricket is a bat-and-ball game played by two teams of eleven players. Despite having several different time-based formats, the objective to accumulate more runs than the opposition remains the same. Fundamentally, cricket is a team game, although a contest between an individual batsman and an individual bowler can be isolated. Generally, the goal of the batting side is to score runs, while bowlers attempt to prevent the scoring of runs. Bowlers have different approaches to restricting the scoring of runs, either by attacking (attempting to take wickets thereby ending a batsman's opportunity to score), or defending (minimising the number of runs conceded).

Traditional metrics for evaluating bowling performance revolve around ratios of three key statistics that are observed on a typical scorecard: wickets taken, runs conceded and overs bowled. The most common measures are bowling average (runs conceded/wickets taken), strike rate (balls bowled/wickets taken) and economy rate (runs conceded/overs bowled).

Lemmer (2012) identifies the recent rise of academic publications surrounding performance measures and prediction methods within cricket. From a bowling perspective, research exists on measurements and comparisons of bowling performance (e.g. Bracewell, 1999; Dey *et. al.*, 2011; Lemmer, 2005; Lemmer, 2006; Kimber, 1993; van Staden, 2008).

Attacking performance for a bowler is determined by either strike rate or bowling average. However, as noted by Bracewell (1999) and van Staden (2008), these metrics are not always defined as a bowler may not have taken a wicket during a match. As cricket is a team game, bowlers work in pairs to create pressure to take wickets for the team. For instance, an individual role in a team's plan could be to restrict scoring in order to create pressure, thereby forcing a batsman to attempt high risk shots to score runs.

The challenge is to recognise an individual's contribution to wicket-taking by the bowling team, when they may not have taken a wicket themselves. A meaningful wicket focused metric for non-wicket takers would be useful for within-game comparisons, creating individual ratings and team ratings (Patel *et. al.*, 2016). The premise of the bowling average is that better bowlers will concede fewer runs per wicket. Importantly, after controlling for batting ability, teams comprised of teams with better performed bowlers are more likely to win (Patel *et. al.*, 2016). Given the value of these metrics as an indication of likely future performance, understanding the contribution of non-wicket takers using the standard bowling ratios in a single game can provide greater insight into potential strategies, team selection and individual performance monitoring.

### REJECT INFERENCE

The requirement to estimate the performance of an individual when data is unavailable is similar to the process for constructing application credit risk scorecards in the financial services sector (Montrichard, 2008). An example of this situation is that of a creditor confronted with the problem of determining whether or not a credit applicant will be able to repay their loan according to a pre-agreed criteria (lump sum or instalments over a fixed time period). Only if the applicant is able to pay back their loan on time and in full are they deemed to have not defaulted, with any breaches in their contract resulting in the applicant being classified as default. Critically, prior to observing this dichotomous outcome, creditors score applicants on their likelihood of default, accepting low-risk applicants. As creditors do not have outcome data on rejected applicants, Reject Inference Methods (RIMs) are implemented to incorporate rejected applicant data and ideally improve model scoring predictive accuracy. Lenders utilise RIMs to salvage rejected applicants to mitigate bias (Verstraeten *et. al*., 2005).

To draw parallels between credit risk and cricket, consider the following comparison. If a bowler takes a wicket then they have a defined strike rate (real outcome). This is compared to bowlers who do not take a wicket leading to an undefined strike rate (unobservable outcome).

### PARCELLING

Most RIMs, and classification algorithms in general, do not output ‘crisp’ categorisation for predictions. Rather, models classifying rejected applicants as default or not default output the ‘fuzzy’ probability of class membership. Parcelling refers to the technique, which can typically be separated from the RIM entirely, that exploits these predicted probabilities for optimal classification (Anderson, 2007) by defining a membership class based on a cut-off which is typically defined by historical data.

## 2. METHODOLOGY

### DEFINITIONS OF DEFAULT AND REJECTION FOR BOWLERS

We consider results from the 2015 IPL, where there were 623 different bowling performances over the season. Data was extracted from Cricinfo (www.espncricinfo.com). Individuals who bowled one over or less during the season were removed, leaving data from 97 individuals.

To align with the typical terminology used within reject inference methods and credit risk, we recast bowlers as having defaulted or not, and whether they are accepted or rejected for reject inference. Both the bowling average and strike rate per game are volatile, since the divisor is typically between 1 and 4 (number of wickets taken in a game), with the results being heavily right skewed. Consequently, the bowlers with strike rates in the top 25% were defined as “non-default”, with all others including non-wicket takers defined as “defaulted”. All non-wicket takers are defined as rejected. The group of accepted wicket taking performances are referred to as *AccApps*, and all bowling performances are referred to as *AllApps*.

To validate our definition of default, we compared default outcome with whether a wicket was taken. The default odds were 2.05 for accepted applicants (0.67/0.33) and 7.41 for those rejected (0.88/0.12). The resulting odds ratio indicates default was 3.62 more likely for rejected than accepted applicants, validating our definitions.

### VARIABLE SELECTION

The extracted scorecard data contained 14 variables. To find the most significant variables for our modelling purposes, we applied a Generalized Boosted Machine Model (GBM) decision-tree based approach. We found GBM to be most suitable for our purposes because of its robust assumption properties (non-parametric, non-linear, ability to handle collinearity and complex interactions). We found the top five variables, listed in order of importance, are: economy rate, number of dot balls bowled (dots), runs conceded, number of boundaries and number of bowler-penalised extras (wides and no-balls).

We tested these variables using a logistic regression on the *AllApps* group, using the data to predict the probability of default. We found that using all of the top five variables led to an Area Under the Curve (AUC), derived from the Receiver Operating Characteristic (ROC), curve of 0.832, while using the top three variables gave an AUC of 0.833. Given the addition of variables does not improve the discrimination of the model, only the top three variables are retained for further analyses.

BASE MODELLING

All records were randomly selected for either the training (70%) or test (30%) dataset. The training set was used to model data structures using the statistical models for each RIM in isolation from the test data. Those trained models were then used to predict outcomes within the test data to produce model performance metrics.

Logistic Regression (LOG) models were used to compare RIMs with the same variables included in each model. Separate LOG models for each RIM output predicted probabilities of "default" on the test data set.

Rejects (non-wicket takers) had their outcome (Strike Rate) imputed by an observed strike rate regression model fitted on all accepted applicant observations (wicket takers) containing the three variables; economy rate, balls and dots. This model is referred to as a non-RIM due to the absence of reject inference.

RIM APPLICATION

All RIMs were programmed in R. RIMs, and corresponding LOG models, were trained under the assumption that real outcomes are unavailable for rejected applicants. The *AccApps* model was substituted as the pre-existing scoring system when this was required for RIM creation.

Parcelling was applied in code blocks for: Fuzzy Parcelling (FUZ; Anderson, 2007; Siddiqi, 2006), No Indeterminates Parcelling (NID) (Anderson, 2007; Siddiqi, 2006), and Proportional Polarised Parcelling (PRP) techniques (Anderson, 2007; Siddiqi, 2006; Thomas, 2009).

FUZ predicts the default probability of rejects using a model that firstly rates the probability of being accepted. Then the rejected applications are sorted by predicted probabilities of acceptance and then classified based on the proportion of the known default rate. Finally, a LOG model is applied on the combined accept and reject datasets.

NID classifies and combines rejects with extreme predicted probabilities, then discards 'indeterminate' rejects. An 'indeterminate' application has an outcome that is unclear. For example, in the financial services sector an individual who is 45 days past due is not technically in default (60 days past due), but is not current either.

PRP adds two copies of rejects to the training data and classifies one copy as default and the other as non-default. The accepted applications are weighted as 1, the first copy of the rejects is weighted as probability of default and the second copy is weighted as the probability of non-default.

There were several other techniques investigated (Anderson, 2007; Finlay, 2012; Siddiqi, 2006; Thomas, 2009; Thomas *et. al*., 2002); However, the best performing techniques from the analyses identified above remain the focus for the remainder of this research.

ASSESSMENT

Three summary statistics that are overall measures of model performance were calculated, including a variation on *Type II Error (*Max5TII), the AUC, and the Mean Absolute Difference (MAD) between predicted probabilities of *AllApps* and the model being tested. Higher AUC values indicate a higher *True Positive Rate* with a lower *False Positive Rate* at all possible operating point values (i.e. all possible probability cut-offs). AUC values of 0.5 and 1 indicate random guess and perfect predictions respectively. In application scoring, an AUC above 0.75 is satisfactory (Anderson, 2007) and 0.80 is "excellent" (Hosmer *et. al.*, 2013), meaning that all LOG models achieved comparable industry best practice in overall performance. Lower MAD values indicate a smaller average difference between estimated default probabilities from the All Applicants (*AllApps*) LOG (i.e. no sampling bias) and the tested LOG in question (i.e. with sampling bias). Max5TII values indicate the number of credit applicants in the test set that were 'accepted' prior to reaching a 5% *Type II Error Rate*. AUC and MAD are the most important measures.

## 3. RESULTS

LOG MODEL

The variables are non-normally distributed, including extreme values and clusters of cases with identical values. Additionally, a linear relationship with the log of the response variable cannot be assumed. Squared Mahalanobis Distances exceeding the $X^2$ critical value of 22.46 ($\alpha$=0.999, df=6) identified multivariate outliers (MVO). Standardised Pearson Residuals (SPR) exceeding critical values of 1.96 ($\alpha$=0.95) identified model outliers. The *AllApps* LOG had 2 MVO and 21 critical SPR, while the *AccApps* LOG had 1 MVO and 14 critical SPR. The records were removed from the training set and the two non-RIM LOGs were retrained. No other MVO or critical SPR of the several hundred were removed. They did not prevent relative performance comparisons across RIMs and comparisons were fairer if each RIM used the data set without further modification.

The two non-RIM LOGs were parameterised and odds ratios with confidence intervals (α=0.95) were calculated along with estimation of coefficients. The estimated coefficients for the two models are similar for most variables, as well as the odds ratios and significance values. The largest difference occurred between the economy rate estimates (1.73 for *AllApps* vs. 0.94 for *AccApps*).

However, there are two areas for concern arising from the accepted applicant (*AccApps*) predicted probabilities which can be observed in Figure 1a. More records were classified as very high-risk. Low risk applicants were overrepresented for parts of the distribution rather than sharing the low density distribution seen in the all applicant (*AllApps*) predicted probabilities.

The scatter plot comparison is shown in Figure 1a. The non-random selection bias is evident with non-linear estimation of default risk across the entire distribution for all records.

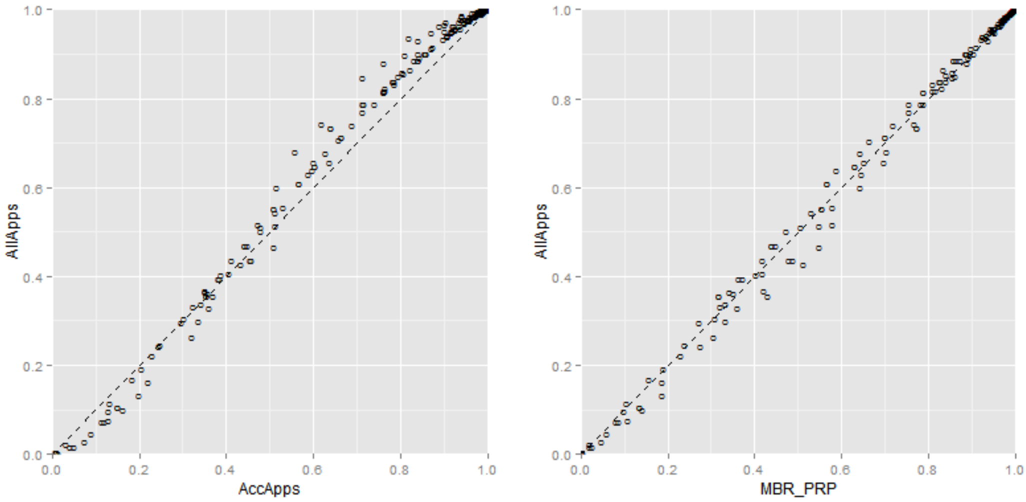



Figure 1a & 1b. Predicted Probabilities Scatter Plot (*AllApps* by *AccApps*) & (*AllApps* by MBR_PRP)

REJECT INFERENCE METHODS

All non-RIM and RIM LOG models were used to predict the applicant population within the test data set using assessment metrics previously defined. MBR_PRP scored the highest AUC and had a reasonably low MAD. It was selected for additional qualitative analysis of predicted probabilities.

This model produced predicted probabilities of default for rejected applicants using Memory Based Reasoning (MBR). In summary, this method applied a nonparametric machine-learning algorithm, *k*-Nearest Neighbours, which estimated probabilities of default for each rejected applicant record using the 30 'nearest neighbours', as determined by Chebyshev distance, out of the accepted applicant records. Probabilities were incorporated using the Proportional Polarised (PRP) parcelling technique. This rank orders rejected applicants by lowest to highest estimated default risk, according to MBR predicted probabilities, then classifies them in proportion to rejected applicant default rate.

The previous comparison between AllApps and AccApps in Figure 1a showed that the non-random sampling bias caused a non-linear representation of the probability of default. Figure 1b shows that MBR_PRP corrected most distortion, especially through high risk regions, and that error is more evenly distributed across all records.

Importantly, the probability of default needs to be converted back to the cricket use case for interpretation. Comparing the probability of default (performance in bottom 75%) with strike rate revealed that a simple linear transformation of multiplying the probabilities by 100 provided a reasonable estimate. This enables the number of wickets per bowler per innings to be estimated, given strike rate is calculated using balls bowled and wickets taken. When a wicket has been taken, this value is used. Where a wicket has not been taken, the inferred wickets are calculated by the inferred strike rate divided by the number of balls bowled. This invariably leads to an estimate of wickets less than 1. Importantly, the season-observed strike rate and the season-adjusted strike rate are correlated (r=0.89), which indicates that undue bias is not introduced.

## 4. DISCUSSION

Using the inferred number of wickets, inferred strike rates and bowling averages can be estimated. If the bowling average is assumed to be an indicator of a bowler's ability, it should be a leading indicator of performance. To test this assumption, the performances of individuals who bowled in six games or more were assessed (n=43). The observed and inferred bowling averages for the last two matches for each qualifying individual were contrasted against the four matches leading up to those two games. 9.1% of the final two matches had observed bowling averages that were undefined, highlighting the usefulness of the inferred estimates.

The inferred bowling average is a better predictor of future performance (using the inferred future bowling average) than the observed bowling average, with a correlation of 0.50 compared with 0.41. Correlations for the observed bowling average, when defined is 0.41 using the inferred result and 0.34 with the observed future result.

These results indicate that two of the three standard bowling measures, bowling average and strike rate, can be used per innings even when a wicket is not taken due to the reject inference method described. Importantly, the method for reject inference identifies the impact of creating pressure on opposition teams by restricting the scoring rate. This further highlights the impact of the bowlers working in pairs to deliver better outcomes for their team.

## 5. CONCLUSION

Lenders use logistic regression models to predict the default risk of credit applicants. A major issue is the non-random sampling bias that occurs due to the systematic rejection of non-credit worthy applications. To remedy this bias, reject inference is used to estimate the performance of a rejected applicant.

This approach was applied to determine the attacking performance of a bowler who did not take a wicket. Standard measures such as bowling average and strike rate are undefined when a bowler has not taken a wicket. Additionally, these metrics do not reflect the contribution a bowler may have had towards wicket taking due to creating pressure by not conceding runs.

To identify relevant attributes for modelling, a Gradient Boosting Machine proved an effective means of optimal variable selection.

The most effective method for reject inference was Memory Based Reasoning with Proportional Polarised parcelling. The model produced predicts that bowlers who deliver a high number of dot balls and a low economy rate have a higher probability of taking wickets.

Importantly, for bowlers with a reasonable workload over the course of a season (bowling in six innings or more), the inferred strike rate is a better predictor of future performances (r=0.50) than the observed strike rate (r=0.41). Consequently, using inferred strike rate and bowling average is a useful mechanism for ranking and monitoring bowling performances. As an attacking metric can be defined for every bowler per innings, this enables holistic bowling performances to be monitored for changes in form and fitness.

## References


Anderson, R. (2007). *The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation*. New York, USA: Oxford University Press.

Anderson, B., Haller, S., Siddiqi, N., Cox, J., & Duling, D. (2010). *Improving Credit Risk Scorecards with Memory-Based Reasoning to Reject Inference with SAS? Enterprise Miner?* SAS Global Forum: SAS Institute Inc.

Bijak, K. & Thomas, L. C. (2012). *Does segmentation always improve model performance in credit scoring?* Expert Systems with Applications, 39(3), 2433–2442.

Bracewell, P. J. (1999). Statistical methods for cricket team selection. Unpublished Masters Thesis, Massey University, Albany Campus, New Zealand.

Bradley, A. P. (1997). *The use of the area under the ROC curve in the evaluation of machine learning algorithms.* Pattern Recognition, 30, 1145–1159.

Croucher, J. (2010). *Player ratings in one-day cricket*. In Proceedings of the fifth Australian conference on mathematics and computers in sport. Sydney University of Technology, Sydney, NSW, 95–106.

Dey, P., K, Ghosh, D., N & Mondal, C., A (2011). A MCMD Approach for Evaluating Bowlers Performance in the IPL. *Journal of Emerging Trends in Computing and Information Sciences*, **2**(11): 564-573.

Finlay, S. (2012). Sample Bias and Reject Inference. In *Credit Scoring, Response Modeling, and Insurance Rating: A Practical Guide to Forecasting Consumer Behavior (2nd ed.).* UK: Palgrave Macmillan.

Friedman, J. H. (2002). Stochastic gradient boosting. Computational Statistics & Data Analysis, 38, 367–378.

Gerber, H., & Sharp, G. (2006). *Selecting a limited overs cricket squad using an integer programming model*. Journal of the Operations Research Society, **5**.

Heckman, J. T. (1976). *The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models*. Annals of Economic and Social Measurement, 5, 475–492.

Hosmer, D.W., Lemeshow S., Sturdivant, R.X. (2013). *Applied Logistic Regression* (3rd ed.). New York: Wiley.

Hothorn, T. (2013). *CRAN Task View: Machine Learning & Statistical Learning*. Retrieved December, 2013, from http://cran.r-project.org/web/views/MachineLearning.html.

Kimber, A. C. (1993). A graphical display for comparing bowlers in cricket. *Teaching Statistics*, 84-86.

Korb, K.B., & Nicholson, A.E. (2011). *Bayesian Artificial Intelligence* (2nd ed., p.183). Boca Raton, Florida, USA: CRC Press.

Lemmer, H. H. (2005). A method for the comparison of the bowling performances of bowlers in a match or a series of matches. *South African Journal for Research, Physical Education and Recreation*, **27**(1), 75-85.

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

Lemmer, H. H. (2012). The single match approach to strike rate adjustments in batting performance measures in cricket. *Journal of Sports Science and Medicine*, **10**, 630-634.

Montrichard, D. (2008). Reject Inference Methodologies in Credit Risk Modeling. Paper ST-160, *Southeast SAS Users Group Annual Meetings*. Retrieved 2013 http://analytics.ncsu.edu/sesug/2008/ST-160.pdf.

Patel, A. K., Bracewell, P.J., & Rooney, S.R. (2016). Team rating optimisation for T20 cricket. *In Proceedings of the thirteenth Australian conference on mathematics and computers in sport*. Under Review.

Ridgeway, G., & Southworth, G. (2013). gbm: Generalized Boosted Regression Models (v2.1). Available from URL http://CRAN.R-project.org/package=gbm.

Sharp, G., Brettenny, W., Gonsalves, J., Lourens, M., & Stretch, R. (2011). *Integer optimisation for the selection of a Twenty20 cricket team*. Journal of the Operational Research Society, **62**(9), 1688–1694.

Siddiqi, N. (2006). *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*. New Jersey, USA: John Wiley & Sons, Inc.

Thomas, L.C. (2009). *Consumer Credit Models: Pricing, Profit, and Portfolios*. New York, USA: Oxford University Press.

Thomas, L.C., Edelman, D.B., & Crook, J.N. (2002). *Credit Scoring and its Applications*. Philadelphia, USA: SIAM.

van Staden, P.J. (2008). *Comparison of cricketers' bowling and batting performances using graphical displays.* Current Science. **96**(6). 764-766.

Verstraeten, G., & Van den Poel, D. (2005). The Impact of Sample Bias on Consumer Credit Scoring Performance and Profitability. Journal of the Operational Research Society, **56**, 981-992.

# REAL TIME PREDICTION OF OPENING BATSMAN DISMISSAL IN LIMITED OVERS CRICKET

Patrick Brown [a,b], Ankit K. Patel [a,b], Paul J. Bracewell [a,b,c]

[a] *DOT loves data*
[b] *Victoria University of Wellington*
[c] *Corresponding author: paul@dotlovesdata.com*

**Abstract**
A number of metrics have emerged to summarise batting performance in cricket, ranging from batting averages to strike rates. However, these metrics are summaries of overall performance and are of limited use in determining in-game tactics.

An expectation of how likely a batsman is to survive each ball faced over the course of an innings can aid the development of more effective strategies to optimise a team's final total. The objective of this research was to formulate a predictive model to calculate the probability of an opening batsman being dismissed in the first innings of a limited overs cricket match. The narrowed focus is designed to eliminate confounding factors such as match state.

Cox proportional hazard models were implemented to consider the potential effect of nine batting performance predictor variables on the ball-by-ball probability of a batsman facing the next ball without getting out.

Through extensive model formation and selection techniques, a contextually and statistically significant Cox model was found. This model was capable of predicting the probability of survival for a particular batsman, given certain conditions. The cumulative number of runs scored, cumulative number of consecutive dot balls faced and cumulative number of balls faced in which less than two runs in four balls had been scored for the batsman were all practically and statistically significant.

The results show that as the magnitude of these three predictor variables increase for a particular batsman, the associated survival probabilities for the batsman either remain constant or decrease on a ball-by-ball basis. Based on One Day International matches played between 8th December 2014 and 8th February 2016, the final model ranked Sangakkara as the most effective batsman at staying in bat for that time period. This method of calculating player rankings is also correlated with ICC ODI rankings, average runs scored and winning.

**Keywords:** Survival, Probability, Cox Proportional Hazards Model

## 1. INTRODUCTION

Cricket is a team sport in which in-game tactics play an important role in success. These tactics can include decisions on batting positions for particular batsman and whether or not an attacking style should be utilized in favour of a more reserved playing style. Such tactics are essential to maximise the team's final total.

Survival analysis is a branch of analysis used to investigate the relationship between one dependent variable, the time until a particular event of interest occurs, and several predictor variables. The survival function refers to the probability that an individual will survive longer than a particular time. The hazard function refers to the rate of failure at a particular time given the individual has survived up until that time. Cox proportional hazard models are commonly used in survival analysis to model the relationship between the hazard function and several predictor variables (Hosmer & Lemeshow, 1999).

The literature revealed two key pieces of academic research that focus on the application of survival analysis techniques to batting in cricket. Kimber and Hansford (1993) were interested in applying methods of survival analysis to batting scores in cricket. In their study, Kimber and Hansford (1993) were interested in a batsman being out as the event and took the number of runs scored to represent time until this event occurred. It was found that fitting a geometric model to the batting average resulted in poor fit and an inconsistent batting average metric. As such, they suggested an alternative batting average that doesn't depend on the geometric assumption and is adjusted for not-out innings. Kimber and Hansford (1993) concluded that a future area of research could be to investigate how additional factors could be combined to illustrate the qualities of particularly strong batsmen. Preston and Thomas (2000) applied aspects of survival analysis to investigate batting strategies in limited overs cricket. In their study, Preston and Thomas (2000) found that batting strategy was driven by the ability to increase the run rate when setting a target, but decreasing the run rate when chasing a target. This study extended on work by Kimber and Hansford (1993) by considering the effect of a number of covariates associated with optimal batting strategy. These included the required run rate and number of wickets lost.

Cox proportional hazards models have been studied extensively. There have been over 25,000 citations of the original paper by Cox since it was published (Bellera et al., 2010). Lane, Looney and Wansley (1986) applied Cox models to the analysis and prediction of bank failure. Nagelkerke, Oosting and Hart (1984) put forward a test statistic to assess the proportional hazards assumption.

This research builds on these techniques. In particular, it addresses the area of future work suggested by Kimber and Hansford (1993) by introducing the effect of a number of potential covariates. It also extends on the work achieved by Preston and Thomas (2000) by incorporating additional covariates that were not considered. In addition, it emphasizes an investigation into batsman dismissal rather than optimal batting strategy. Specifically, a model has been developed consisting of a variety of predictor variables capable of predicting the probability of opening batsman being dismissed in the first innings of a limited overs cricket match.

## 2. METHODS

DATA COLLECTION

Ball-by-ball data was extracted from Cricinfo commentary (www.espncricinfo.com) for One Day International (ODI) cricket matches contested between 8th December 2014 and 8th February 2016. For each ball faced, data consisting of a number of variables were collected. These included the match, innings and player identifiers, over and ball numbers, bowler and batsman-facing metrics and outcomes from that ball. Those outcomes included if there was a dismissal, number of runs scored and number of extras (only wides and no balls are considered due to the audit trail within the data extract). Matching this transactional information with the scorecard data enabled batting position to be established. Data collection was restricted to within-game events.

DATA MANAGEMENT

Once data had been collected, data management proceeded by removing data not associated with the first innings of games and data associated with batsmen in positions three or higher. The narrowed focus was designed to eliminate confounding factors such as match state. Variables that could potentially have an effect on the prediction of the probability of a batsman being dismissed and could be calculated from the original variables were then considered. The cumulative number of runs scored for the batsman and for the team were two factors. Similarly, the cumulative number of balls faced by the batsman and the team were also of interest. Other factors considered included batsman strike rate, run rate and batting average as well as dot ball and consecutive dot ball effects. Another factor was the number of balls faced by the batsman in which less than two runs in four balls had been scored. Calculation and incorporation of these variables followed on from data manipulation (see results).

MODEL FORMATION AND SELECTION

This research utilized the Cox proportional hazard modelling technique. A survival object was created and taken to represent the response variable in a Cox proportional hazards model. This consisted of a particular event and the time taken to that event, in this case the event being a batsman getting out. It made sense to subsequently use the total number of balls faced by the batsman to represent the time to that event.

This research progressed by fitting a large number of Cox models, with each individual model consisting of a different combination of predictors. For each fitted model, three selection criteria were favoured over conventional model selection techniques such as AIC. Model selection was important in order to narrow down a set of candidate models. To meet the first criteria, the estimated model coefficients had to make practical sense. For example, an increase in predictors such as cumulative runs scored was expected to decrease the likelihood of a batsman surviving the next ball. For these predictors, the first criteria was met if the corresponding estimated coefficient was negative. Similarly, an increase in resource availability was expected to increase the likelihood of a batsman surviving the next ball. In this situation, the first criteria was met if the corresponding estimated coefficient was positive. To meet the second criteria, the predictors had to be statistically significant. If these first two criteria were met, the third condition specified that the probability of batsman survival either remained unchanged or decreased on a ball-by-ball basis**.** Models that were close to meeting either of the first two criteria were also kept in the candidate set. The reason behind this was to be able to assess whether insignificant model predictors would become practically and statistically significant if transformations were applied.

There are two assumptions that the Cox model relies on. The first states that the effect of each predictor is linear in the log hazard function. The second key assumption states that the ratio of the hazard function for two individuals with different sets of covariates does not depend on time, suggesting the hazards are proportional. This research utilized further unconventional methodology to select a final model to be used for prediction. Once the candidate models had been narrowed down, these assumptions were examined to check whether each model showed any evidence of violation. Through evaluation of these assumptions, the candidate set of models was narrowed down further. From this reduced candidate set of models, the final model was selected to be the one believed to incorporate the highest variety of data associated with each batsman. The variables in the final model were: the cumulative number of runs scored, cumulative number of consecutive dot balls faced and cumulative number of balls faced in which less than two runs in four balls had been scored.

PREDICTION

The ball-by-ball survival probabilities for all 43 opening batsman considered were calculated using:

$$\log(p/(1-p)) = \exp(\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3) \qquad (1)$$

where $p$ represents the probability of survival and $\beta$ represents the weights for each attribute, $x$. The survival probabilities for each opening batsman in each game were plotted to produce survival curves. Figure 1 was constructed to illustrate the ball-by-ball survival probabilities for Guptill and McCullum in the ODI game between New Zealand and Sri Lanka on 14th February 2015. Figure 1 also illustrates the ball-by-ball survival probabilities for Warner and Finch in the ODI game between Australia and England on 14th February 2015. The area under each survival curve and the total area under all curves for each batsman were calculated. To account for the differing number of games played by each batsman, the average area under the curve for each batsman was computed. This was used as a metric for batsman comparison purposes. Another metric, the wins-to-games ratio for each batsman was also calculated. An area-to-games ratio metric was regressed against the wins-to-games ratio to assess for a relationship between the likelihood of a batsman getting out and the number of wins they achieve.

## 3. RESULTS

The final model capable of predicting the probability of survival for a particular batsman included three predictor variables. These consisted of the cumulative number of runs scored, cumulative number of consecutive dot balls faced and cumulative number of balls faced in which less than two runs in four balls had been scored for the batsman. Originally, only the first two variables were practically and statistically significant. The estimated coefficients for these two predictors made practical sense and associated p values from a z-test were less than the significance threshold of 0.05. Through evaluation of the linearity assumption of a Cox model, all three predictor variables showed evidence of non-linearity. To correct for this assumption, a square root transformation was applied to the cumulative number of runs predictor and a log transformation was applied to the cumulative number of consecutive dot balls predictor. A log transformation was also applied to the predictor consisting of the cumulative number of balls faced conditional on less than two runs scored in four balls. Once non-linearity was corrected for, the cumulative number of balls faced in which less than two runs in four balls had been scored for the batsman became practically and statistically significant. Evaluation of the proportionality assumption resulted in no statistical evidence of any violation towards this assumption.

Figure 1 illustrates a selection of the results from this model. This model suggests that in the ODI between New Zealand and Sri Lanka, Guptill had higher survival probabilities than McCullum. In the ODI match between Australia and England, Finch had higher survival probabilities than Warner. These results are likely to be of interest to those associated with New Zealand and Australian cricket.

Of interest is the apparently different roles adopted by each pair. Given the survival properties, the results imply that McCullum and Warner have opted for a higher risk strategy, whilst Guptill and Finch have been more conservative. Optimal partnership strategies could be derived through further research.

As an extension to this research, the final model and associated metric calculations and comparisons could be applied to the rest of the batting team in the first innings and to the opposition batsmen in the second innings.

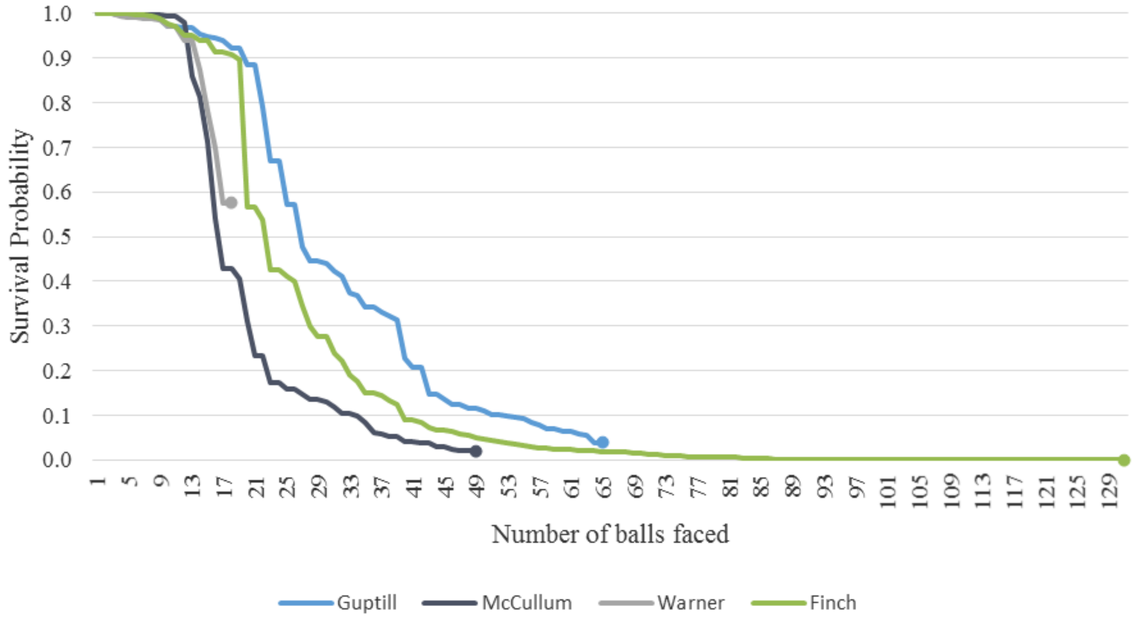



Figure 1: Survival probabilities from 14th Feb 2015 for Guptill and McCullum in the ODI game between New Zealand and Sri Lanka and for Warner and Finch in the ODI game between Australia and England

As the average number of observations per batsmen is 3.16, results for the 43 batsmen were aggregated into groups of 5 according to their rank ordering based on the average area under the curve. This enabled the average area statistic to be assessed as a meaningful measure of performance. The assumption is that a higher area indicates longer periods of time spent at the crease and is therefore indicative of better performances.

Firstly, the ODI ICC batting ranking for each batsmen was obtained (www.relianceiccrankings.com). The ranking used was either that following their last international (for retired players) or the ranking as of 8th February 2016. The average number of runs scored for the cohort was also calculated. Each rank-ordered cohort consists of 5 batsmen. Figure 2 compares the average area under the survival curve against the average ICC ranking per player. The size of the bubble is based on the average number of runs scored per cohort, with that average used to annotate the graph. The top cohort contained Sangakkara, Samuels, Haque, Porterfield and Thirimanne.

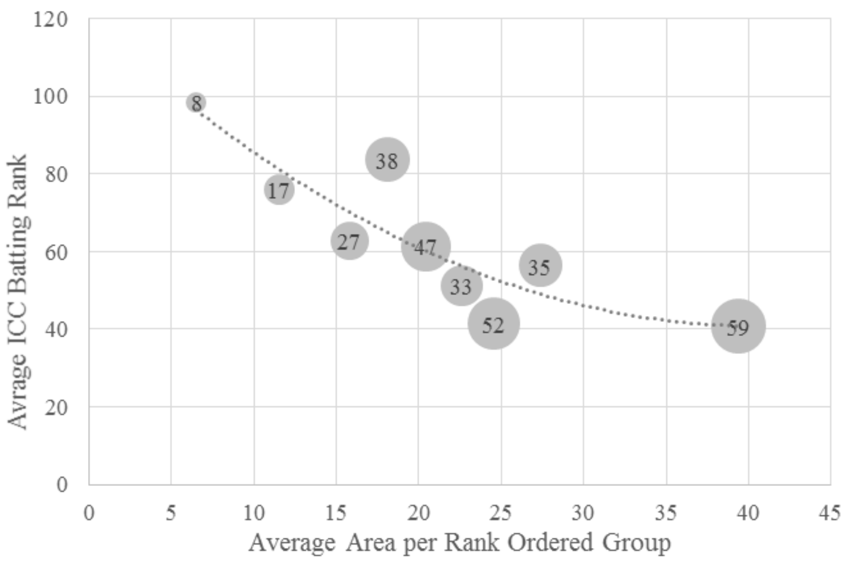



Figure 2: Average area for rank ordered cohorts compared with ODI ICC batting ranking and average number of runs scored for the observed time frame (within bubble)

There is clearly a strong relationship between average area and the average ICC rankings. The correlation between the square root of the average area and the square root of the average rankings is 0.91, while the square root of the average area and the square root of the average runs is -0.87. This result is not surprising as in order to score runs, an opening batsman needs to occupy the crease.

The proportion of games won within each cohort is also explored. The lowest ranked cohort won 0 of 6 games. As a consequence this outlier has high leverage and inflates the strength of the relationship between the average area and winning. After removing that observation, an $R^2$ of 0.17 is obtained. The direction and magnitude of this statistic indicates that opening batsmen can have a strong impact upon match results by batting for sustained periods of time whilst accumulating runs.

## 4. DISCUSSION AND CONCLUSION

Through development of a model capable of calculating the ball by ball probability of a batsman getting out, we found a different perspective on the assessment of batting performance. The survival curve that is generated enables the risk of different individual batting strategies to be assessed. For instance, comparing the performance of New Zealand opening pair Brendon McCullum and Martin Guptill against Sri Lanka on 14th February 2015, revealed that McCullum opted for a higher risk strategy on route to 65 from 49 balls (10×4 runs, 1×6 runs) compared with Guptill's 49 from 62 (5×1 runs, 0×6 runs).

Using the average area under the survival curve as a method for ranking opening batsmen, the final model suggests that of the 43 opening batsmen considered, Sri Lankan great, Kumar Sangakkara is the most effective at occupying the crease. When Sangakkara retired immediately after the 2015 Cricket World Cup, he was ranked number 2 in the world. This statistic extends understanding of batting performance from batting totals and strike rates to encapsulate the risk and in-game strategies adopted by batsmen and teams. As this is inherently linked the manner in which a batsman approaches an innings, it is suitable for further research for optimising the output of batting partnerships. Furthermore, this type of analysis may be useful for scouting youth talent.

Given the increased interest in short forms of the game, particularly T20, statistics that demonstrate the positive influence that occupying the crease for opening batsmen has on match outcomes is useful for player development, selection and in-game strategies. Extending this research to T20 matches is clearly of interest.

Importantly, this model extends understanding of batting performance from considering just batting totals and strike rates to encapsulating risk and in-game strategies adopted by batsmen and teams. As this is inherently linked to actual team approaches, it is suitable for further research for optimising the output of batting partnerships.

The availability of machine readable access of ball-by-ball data enables deeper understanding and derivation of in-game strategies. This research has highlighted the use of survival analysis as a suitable technique for investigating cricket and identified several areas of future research.

### References

Bellera, C., MacGrogan, G., Debled, M., de Lara, C., Brouste, V., & Mathoulin-Pélissier, S. (2010). Variables with time-varying effects and the Cox model: Some statistical concepts illustrated with a prognostic factor study in breast cancer. *BMC Med Res Methodol*, **10**(1), 20. http://dx.doi.org/10.1186/1471-2288-10-20

Crowder, M. (2001). *Classical competing risks*. Boca Raton: Chapman & Hall/CRC.

Gamel, J., Vogel, R., Valagussa, P., & Bonadonna, G. (1994). Parametric survival analysis of adjuvant therapy for stage II breast cancer. *Cancer*, **74**(9), 2483-2490.

Hosmer, D., & Lemeshow, S. (1999). *Applied survival analysis*. New York: Wiley.

Kimber, A., & Hansford, A. (1993). A Statistical Analysis of Batting in Cricket. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, **156**(3), 443.

Lane, W., Looney, S., & Wansley, J. (1986). An Application of the Cox Proportional Hazards model to Bank Failure. *Journal of Banking & Finance*, **10**(4), 511-531. http://dx.doi.org/10.1016/s0378-4266(86)80003-6

Nagelkerke, N., Oosting, J., & Hart, A. (1984). A Simple Test for Goodness of Fit of Cox's Proportional Hazards Model. *Biometrics*, **40**(2), 483.

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

# CRICKET AS LIFE AND DEATH

Bernard Kachoyan [a,c] , Marc West [b]
[a] *University of New South Wales*
[b] *Defence Science and Technology Group*
[c] *Corresponding author:* bernkach@outlook.com

## Abstract

The scoring record of a batsman in cricket is usually described in terms of batting averages and other bald statistics such as the number of centuries scored. These simple single figure measures do not fully capture the richness of a typical batsman's career, nor do they address more complex questions such as what are the chances that a batsman will get out before a certain score. A more detailed and informative way of expressing a batsman's career can be derived by considering the experience of a batsman's innings as a lifespan: when you go out to bat you are "born", you "live" for a number of runs and when you get out you "die". The list of scores can then be analysed in terms of a survival function, as used in medical and engineering applications, that is, the probability that something (a product, a patient, a system) will survive beyond a specified time. When a batsman remains Not Out (NO) at the end of an innings they do not 'die', but they have stopped 'living', and this needs to be accounted for in the methodology. This is equivalent to the subject leaving the experimental sample pool after being observed for a while. A standard treatment of this type of data is through the use of Kaplan-Meier, or Product Limit, estimators. This paper will show how a batting record can be expressed in terms of survival function curves using Kaplan-Meier estimation techniques. It will be shown how this can be used to visually present and compare the histories of different batsmen and can lead to other interesting insights.

**Keywords:** Cricket, batting, Kaplan-Meier, Product Limit Estimators, survival functions

## 1. INTRODUCTION

Ever thought of batting as a life and death struggle against hostile forces? The career scoring record of a batsman in cricket is usually described in terms of batting averages and other bald statistics such as the number of centuries scored. These simple single figure measures do not fully capture the richness of a typical batsman's career, nor do they address more complex questions such as what are the chances that a batsman will get out before a certain score.

The experience of a batsman's innings can be described in terms of a lifespan: when you go out to bat you are "born", you live for a number of runs, and when you are dismissed you "die". The series of these innings are thus a sample of such lifespans.

Thought of in this way, the innings can be analysed in terms of survival functions (Ibrahim (2005)): the probability of survival as a function of time, or the probability that the survival time of an individual exceeds a certain value, given by $S(t) = P(T > t)$, where T is the survival time.

Here we will apply those ideas to batting in cricket. In this application "time" is interpreted as a number of runs scored. It is not entirely a new idea to apply these techniques to cricket (e.g. Danaher (1989), Kimber and Hansford (1995)). This paper extends these ideas by considering the whole curve not just individual metrics. It also makes use of the complete cricket databases and computing power now available.

In this application, it is important to properly consider what happens when a batsman remains not out at the end of an innings. Cricket analysts have considered many ways of treating these events (Cohen (2002)), but they are a standard part of a survival interpretation. In these cases, the batsmen do not 'die', but do stop 'living', equivalent to a subject being observed for a while, then leaving the sample pool before the final outcome is achieved. This is termed *censored* data.

## 2. ESTIMATION OF SURVIVAL FUNCTIONS

These types of problems are commonly addressed using Kaplan-Meier (KM) estimators, more commonly known as the Product Limit Estimator (PLE) (Kaplan and Meier (1958), Klein and Moeschberger (2003), Ibrahim (2005)). An important property of the PLE is that it is non-parametric. The PLE only uses the data to generate an estimate of the "true" underlying survival function. Another important advantage of the PLE method is that it can take account of censored data in a relatively straightforward manner. The PLE is the unbiased maximum likelihood estimator of the survival function of the underlying population.

The formulation of the PLE is based on the conditional probability that an individual dies in the time interval from $t_i$ to $t_{i+1}$, given survival up to time $t_i$ is estimated as $d_i / n_i$ where $d_i$ is the number who die at time $t_i$, and $n_i$ is the number alive just before time $t_i$, including those who will die at time $t_i$

The formulation is intuitive and fairly easy to calculate, especially when expressed recursively:

$$S(t_j \leq t < t_{j+1}) = S(t_{j-1} \leq t < t_j) \frac{n_j - d_j}{n_j} \quad (1)$$

Without censoring, the PLE at any given time is simply the number still alive divided by the number originally in the sample. The use of the recursive formulation (1) makes it easier to consider censored data. The PLE survival estimator is piecewise constant with discontinuities at the times of death[1]. The PLE approaches the true survival function for that population as the sample size increases.

The PLE curves provide a visual depiction of all of the raw data, often including explicitly marking the times of censoring, and give a sense of the underlying probability model to guide or even obviate the need for further analysis.

There are several drawbacks in this estimator, however, some of which will be seen in the examples below. Firstly, the vertical drop at specific times is drawn from the data, and do not indicate specific "danger times". Also the probability of surviving each "danger time" depends only on the number of items at risk at that time, not the specific time of censoring. The PLE also gives no prediction of performance beyond the largest data point if the highest score is censored. Finally, the reduction of the sample at large values means the effect of each individual failure on the size of the step-down increases and the accuracy of the estimate decreases at long times,.

## 3. APPLICATION TO CRICKET

In applying the PLE to cricket, a batsman's death means being dismissed, being censored means completing the innings before being dismissed (remaining NO), and time is interpreted as number of runs scored. The batting survival function is then probability that a batsman will score > $x$ runs or $S(x) = P(X > x)$ .The time intervals are equivalent to the run intervals between dismissals ($x_j$ = #runs of $j$ th dismissal).

Unlike many applications, cricketers can, and often do, get 0 runs (or ducks), so the survival functions will not in general start from unity at $x = 0$. In addition, unlike true time, the number of runs is a discrete quantity not a continuous one. This is not a problem for the PLE analysis, and it is often the case in practice that the experimental time is discrete (for example, patients monitored at regular intervals).

We have initially restricted our analysis to long form matches as these are the most conducive to high scoring innings, have a longer history for comparison purposes and to bound the scope of the analysis. We have used the cricket database from cricinfo (2016) and cricketarchive (2016) for the data in this paper. These are extremely rich datasets which can be mined from many different angles. To illustrate the survival analysis concepts, consider the batting statistics of Steve Waugh (SW) and Sachin Tendulkar (ST). Both are highly rated with a similar career batting average[2] (51.5 for SW and 53.8 for ST) and have a large number of innings, that is, a relatively large sample size. SW has a much larger ratio of NOs (18%) than ST (10%)[3], so we can highlight the effect of these on the survival estimates.

Without the consideration of the censored data (the NOs[4]), the PLE survival curve simply reverts to the percentage of scores less than or equal to a certain number of runs. This is shown as 'uncensored' in Figure 1. Including the NOs in the formulation, we get the curves marked 'censored' in Figure 1. As expected, the survival rates increase, particularly in Waugh's case from noticeably below Tendulkar's to closely mirroring it. The relatively large number of NOs (11) for SW at around 120-160 runs are clearly shown in the survival curve and strongly affect the behaviour of his survival function at higher scores.

The integral of the survival curve gives the expected lifetime. This measure has been used by some authors to get a better measure of batsmen ranking than traditional averages (for example Danaher (1989)). Discussion of this type of analysis is outside the scope of this paper, however.

[1] Or whatever event is being monitored

[2] Defined in the traditional sense of number of runs scored / number of dismissals

[3] In fact he is renowned for his batting with lower order batsmen (see http://thecricketcouch.com/blog/2013/01/02/playing-with-the-tail/

[4] Technically, a batsman can also be "Retired - not out", having stopped batting because of injury, illness or other cause. We include these together as NO as they are mathematically equivalent.

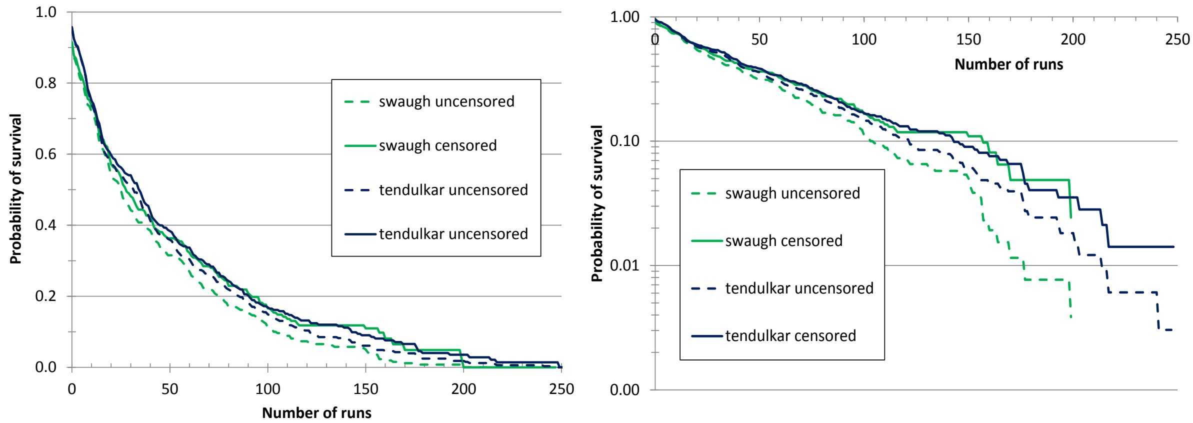



Figure 1 Product Limit Estimator survival function for Steve Waugh (SW) and Sachin Tendulkar (ST), plotted on linear (left) and logarithmic (right) scales.

## 4. COMPARISON BETWEEN BATSMEN

Confidence intervals can be placed on the derived curves using the so-called Greenwood formula (Greenwood (1926). These are less accurate in the tail of the curves, where by definition the sample size is smallest. Unfortunately, as we will see it is often in the tails of the curve where the distinctions between batsmen are found. There are other formulations (Klein and Moeschberger (2003)) which if anything tend to increase the size of the intervals.

To illustrate how the use this formulation in comparing batsmen, we now consider the survival curves of two specialist batsmen with significantly, in a cricket fan sense, different averages. We choose Steve and Mark Waugh (SW and MW), not just because they are brothers but because the latter's batting average, while quite respectable at 41.8 does not put him in the top ranks amongst cricket enthusiasts. The curves, see Figure 2, show similarities at low scores with MW following constant hazard (log-linear) behaviour very similar to SW at scores up to about 50 then an increasing drop off thereafter. From these curves we can surmise that MW was as reliable a batsman as SW in getting to about 50 or so runs but then fell away in performance.[5]

We have also shown in Figure 2 the confidence intervals for these players using the Greenwood formula with a 95% confidence level further emphasising the separation of the curves for larger scores.

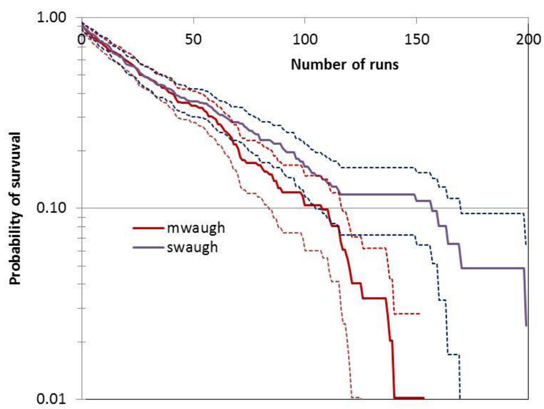



Figure 2 Estimated survival functions for Steve Waugh (SW) and Mark Waugh (MW). 95% confidence limits are shown as dotted lines.

A number of ways of statistically comparing the whole survival curves exist. In this paper we shall use the unweighted Log-rank test. The log-rank test for difference in survival between two groups is the most powerful test when the proportional hazards assumption holds (Klein and Moeschberger (2003)), which as we shall see is commonly assumed to hold for this data and is a reasonable assumption for at least parts of the batting survival functions. It is clear from the extent of the confidence limits in Figure 2 that we might be hard pressed to prove



[5] This is evidence supporting cricketing folklore that MW lost concentration after a period of batting.

statistically significant differences between the SW and ST survival curves and this is confirmed with the log-rank test, although many observers rate Tendulkar as one of the best batsman of all time. Log-rank analysis of SW and MW shows no statistical difference (just) at the 95% level using the whole dataset but strong significant differences (>99%) if consideration is restricted to greater than 100 runs.

**5. BATTING HAZARD FUNCTION**

The instantaneous *hazard function* (or failure rate) *h(t)* is defined from the reliability/survival function, S(t), and is given by $h(t) = - S'(t)/S(t)$. The hazard function is a *conditional* probability of the failure density function S'(t), conditioned that failure has not occurred at time *t*.

If the failure distribution is exponential then the failure rate is a constant and vice versa. This is known as a memoryless property. The discrete equivalent of this distribution is the Geometric, which strictly speaking should be used for the discrete data being considered here. The Geometric distribution approaches the exponential distribution when the sample size is large and the probability of each change is small. Cohen (2002) shows how a Geometric distribution can be derived from assuming, inter alia, that both the probability an innings ends with each ball faced and the probability batsman makes a scoring shot with each ball faced is a constant, thus making ball to ball independence explicit. Whether batting follows a memoryless distribution or not is a profound concept for cricket enthusiasts. It implies that no matter what score you are on you have the same probability of getting out. It also influences the notion of batting form - is there such a thing if no matter what score you are on, you have the same chance of being dismissed?

A pioneering study by Wood (1945) indicated that a Geometric distribution was broadly representative of the batting data he considered. This conclusion has been disputed in later studies (e.g. Kimber and Hansford (1995)), but many subsequent analyses have reinforced its applicability at least in some instances (e.g. Barr and der Honert (1997), Bracewell and Ruggiero (2009), Bailey and Clarke (2004)). One common characteristic is the tendency of Geometric distribution to underestimate the number of 0s (or *ducks*) and overestimate the number of 100+ scores.

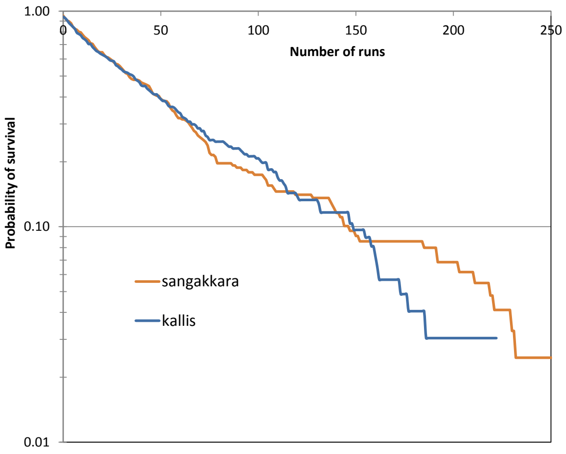
A simple test of whether the batting statistics follow a memoryless distribution is whether the logarithmic plots of the PLE are straight lines with the slope of the line of best fit giving the failure rate. A visual examination of Figure 1 shows that the linear approximation may indeed be valid until the high scoring regime, and excepting very low scores.

For scores less than 100, the failure rate for SW and ST is about 0.016, or 1.6% per run, with very good correlation to linear ($R^2$ = 0.99)[6]. Scores over 100 show a decrease in linearity ($R^2$ for SW = 0.86) and for ST an increase in hazard rate (0.0195) which could be simply a result of a decreased sample size and a concentration of NOs in this regime, but also reflects these batsmen behaving differently once they have reached their century.

Figure 3 shows the PLE of Jacques Kallis (JK) of South Africa and Kumar Sangakkara (KS) of Sri Lanka. These are recent batsmen of similar high averages (55.4 vs 57.4) and relatively high number of innings. Both batsmen show very similar behaviour up to about 75 runs, but beyond that more differences are apparent than between SW and ST. Like SW, JK has a large number of NOs compared to KS (40 vs. 17), almost 25% of these being at scores of 130 or more. On the other hand KS has a greater number of very high scores (11 vs 2 of >200). Both of these can be clearly seen through observation of Figure 3. JK and KS also appear to show some determination to reach 100 as can be seen by decreasing failure rate on approaching this milestone. A log-rank analysis does not find statistically significant difference between the two players.



[6] Batting data has a chance of failure at exactly zero; hence the LOBFs have a non-zero intercept.



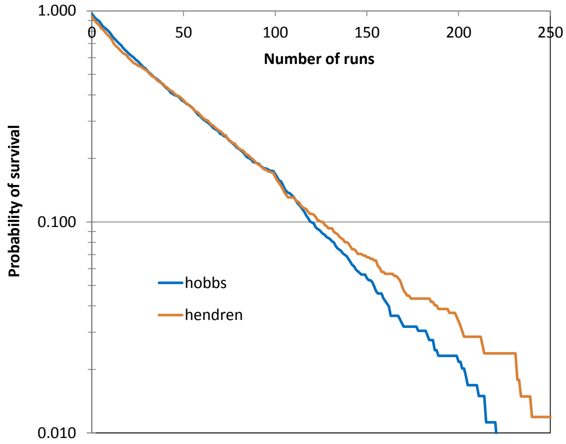




Figure 3 PLE survival function for Kumar Sangakkara (KS) and Jacques Kallis (JK) (left); and Jack Hobbs (JH) and Patsy Hendren (PH) (right).

In order to show the effect of high sample size on the PLE curves, Figure 3 also shows the first class careers of English players Jack Hobbs (JH) and Patsy Hendren (PH). These were chosen due to the extremely high number of innings (>1300), high and similar batting averages (50.7 and 50.8 respectively) and overlapping playing periods. The curves are smoother than those previously discussed yet still show the piecewise linearity and effects of higher NO scores beyond about 200 runs. The similarity of JH and PH survival up until 100 runs is remarkable, as is the obvious deviation for > 100. For PH the hazard rate remains very similar across most of the data set whereas for JH a much better fit to a constant hazard is gained if we consider <100 and >100 (failure rate 0.017 vs 0.023) as separate regimes. The score less than 100 are also a better fit to linearity ($R^2$ of 0.998 vs 0.985). A log-rank analysis confirm no statistical difference if the whole dataset is included but significant difference (>95%) appear if more than 100 runs is only considered. This is evidence supporting the anecdotal statement of Wood (1945) that JH would get himself out after scoring 100 as he had already given the crowd their entertainment. The detailed difference between these curves for players with very similar bulk statistics is indicative of the ability of this type of analysis to tell a broader story.

## 6. APPLICATION TO ALL PLAYER STATISTICS

Modern computing allows us to apply the PLE technique to large data sets, so that we are able to compare not only players, but teams, batting positions, left and right handers, how out etc. Using every Test innings ever played, the technique allows us to investigate accepted cricketing wisdom, such as the best batsman bats at number 3 in the batting order, and that batting in the 4th innings of a Test is much harder than the others. These are illustrated in Figure 4.

The much larger data sets, a total of about 76,000 innings, give smoother survival curves than for individual players, however the discontinuities in the curves are still evident at the tails. Note that each curve is constructed from different distributions representing different batsmen; hence at best reflect a mixed exponential distribution, not an exponential per se. Figure 4 shows that batting positions 3 and 4 indeed have the best performance, but they have very similar survival curves[7]. Interestingly the opener listed as number 1 has a higher survival rate than the opener listed at 2. It may be then that the better batsman generally faces up first. The curve for number 5 also clearly shows a reduction in hazard rate as these batsmen approach the 100 and 200 milestones.

In terms of which innings is the easiest to bat in, Figure 4 suggests that the scores in innings 3 and 4 are indeed less than those in 1 or 2. This could be due to the difficulty of batting on a deteriorating pitch, but may also reflect the less open-ended nature of these innings compared to the first innings[8] (see also Borooah and Mangan (2010)). The greater number of NOs in innings 4 can also clearly be seen.

[7] The high sample size means that there is a quite high power of the log-rank to detect significance, but none was found between these.

[8] That is teams can either run out of playing time or stop batting once they have passed a required total.

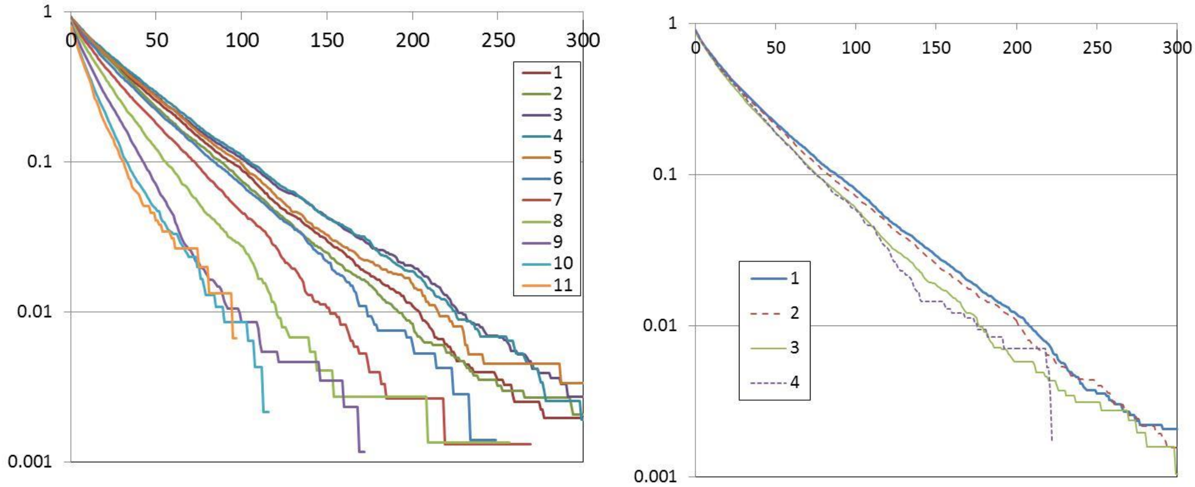



Figure 4  PLE survival function for all batsmen in Test matches split by batting position (left) and innings number (right)

## 7. CONCLUSION

We have shown how a batsman's total career can be illustrated through the use of survival curves. These can also be used to compare the characteristics of batsmen in a way that measures like average or percentiles struggle to achieve.  In particular the performance of batsmen at different parts of their innings can be inferred.

Looking at the complete survival curves provides an explanation for the mixed success of tests for memoryless distributions. For many batsmen this provides a very good fit, but only to parts of their performance curve, generally at scores of less than 100, sometimes considerably less. In addition some batsmen's behaviour can be split into different regimes, low and high scoring.

Because of the richness of the data set this type of analysis can also be used to explore various global assumptions about batting and batsmen in cricket. Conversely, a single batsman's career can easily be split into phases or various opponents etc.

This analysis has also illustrated the drawback of applying strict statistical tests in areas where their applicability is not clear, particularly with the small, in statistical terms, sample sizes of a single batsmen's career.

## References

Bailey, M. J., & Clarke, S.R. (2004). Market inefficiencies in player head to head betting on the 2003 cricket world cup. Cricket World Cup. Economics, Management and Optimisation in Sports (P. Pardalos & S. Butenko eds.). pp. 185-201. New York: Springer.

Borooah V.K. and Mangan J. (2010) The "Bradman Class": An Exploration of Some Issues in the Evaluation of Batsmen for Test Matches, 1877-2006. Journal of Quantitative Analysis in Sports; 6(3):14-14.

Bracewell, Paul J. and Ruggiero, Katya (2009) "A Parametric Control Chart for Monitoring Individual Batting Performances in Cricket," Journal of Quantitative Analysis in Sports: 5(3), Article 5.

Cohen G.L. (2002) Cricketing Chances. Proceedings of the sixth Australian conference on mathematics and computers in sport. Cohen G and Langrtry eds. 1-3 July 2002. Pp 1-13

Cricinfo (2016) Retrieved from [http://stats.espncricinfo.com/ci/engine/stats/index.html](http://stats.espncricinfo.com/ci/engine/stats/index.html)

Cricketarchive (2016) Retrieved from http://cricketarchive.com/

Danaher J. D.  (1989). Estimating a cricketer's batting average using the product limit estimator. The New Zealand Statistician, 24(1), 2-5.

Greenwood, M. (1926). "The natural duration of cancer". Reports on Public Health and Medical Subjects (London: Her Majesty's Stationery Office) 33: 1–26.

Ibrahim. J.G. (2005) Applied Survival Analysis. The 21st Annual Summer Workshop of the North-eastern Illinois Chapter of the American Statistical Association. [http://www.amstat.org/chapters/northeasternillinois/pastevents/summer05.htm](http://www.amstat.org/chapters/northeasternillinois/pastevents/summer05.htm)

Kaplan E.L. Meier P. (1958) Non-parametric estimation from Incomplete Observations. *American Statistical Association Journal*. Vol. 53, No. 28, 457-481

Kimber A.C. and Hansford A.R. (1995) A statistical analysis of batting in cricket. *J. royal statistical society* A, 156 443-455.

Klein, J.P. & Moeschberger M.L. (2003) Survival Analysis: Techniques for Censored and Truncated Data Springer 2nd Ed.

Wood G.H. (1945) Cricket scores and geometrical progression. J. Royal Statistical Society, 108, 12-22

# TEAM RATING OPTIMISATION FOR T20 CRICKET

Ankit K. Patel [a,b], Paul J. Bracewell [a,b,c], Samuel J. Rooney [a]

[a] *DOT loves data*
[b] *Victoria University of Wellington*
[c] *Corresponding author:* paul@dotlovesdata.com

## Abstract

Cricket is an ideal sport to isolate individual team member contribution with respect to winning. This is due to the volume of digital data available, combined with the relatively isolated nature of the batsman versus bowler contest observed per ball.

As cricket, like many other sports, is reliant on the contribution of individuals and their interactions, there are fluctuations in match outcomes. Understanding the root causes of this variation can help interested parties derive insight into team success and potential strategies for optimising performance.

Understanding the individual dynamic within the team setting can lead to improved team ratings. The objective of this research was to develop a roster-based optimisation system for limited overs cricket by deriving a team rating as a combination of individual ratings. The intent was to build an adaptive optimisation system that selects a cricket team of 11 players from a list of available players, such that the optimal team produces the greatest team rating.

The attributes used to define the individual ratings are based on the statistical significance and practical contribution to winning. An adaptive system was used to create the individual ratings using a modified version of a Product Weighted Measure. The weights for this system were created using a combination of a Random Forest and Analytical Hierarchical Process.

The underlying framework of this system was validated by demonstrating an increase in the accuracy of predicted match outcomes compared to other established ranking methods for cricket teams. For the 2015 IPL, this approach outperformed published subjective assessments by 20% and an implementation of an objective calculation by 13%.

The results show that cricket team ratings based on the aggregation of individual playing ratings with attributes weighted towards winning limited over matches are superior to ratings based on summaries of team performances and match outcomes.

**Keywords**: Adaptive System, Product Weighted Measure, Analytical Hierarchical Process

## 1. INTRODUCTION

The growth of sport analytics and the need for meaningful sport related statistics has emerged in recent decades due to the popularity of professional sport as live and televised entertainment. This has led to increased investment in players and teams. The rise in player salaries and salary caps over the last 25 years provide ample evidence of the growth of sport analytics, with investors, franchises, clubs and other stakeholders wanting to determine the true value of their investment. For example, in the National Football League (NFL) there has been an increase of approximately 950% in player salaries since the 1980's, and an increase of 288% in salary cap since 1994 (Vroom, 2012). With global sports revenue estimated to grow by US$145.3 billion over the 2010-2015 period (Fenez & Clark, 2011) and the large investment of resources and the stakes involved, coaches and managerial staff cannot solely rely on subjective views and personal beliefs to make team and player selection decisions.

The explosion in the sporting industry in terms of popularity and revenue is evident in cricket. Cricket has seen huge global growth in revenue in recent years, and transformed into a sporting juggernaut due to the advent of T20 cricket. The Economist reported that global cricket will generate total revenues of approximately $2.5 billion over the period 2014-2022.

Given the myriad of numerical data generated by sports, it is paramount that meaningful information is extracted from the data. There is a breadth of academic literature applying various statistical techniques to myriad sports. For example, Di Salvo *et. al.* (2010) utilised discriminant analysis to identify performance metrics that significantly distinguish between winning, losing and drawing team in the Europe Champions League. Annis *et. al.* (2005) claimed that traditional win/loss and points scored ranking models applied to American Football fail to produce satisfactory rankings. The study therefore developed a hybrid paired comparison model which outperformed competitor models, producing robust results under model misspecification. Further, a modified least squares ranking procedure was developed in by Harville (2003) to rank division 1 American men's college basketball teams using game outcomes. The results showed that the predictive accuracy of the modified least squares (76.3%) method outperformed that of the basic least squares (74.2%).

Cricket has recently seen an exponential rise in the use of statistics to make informed and strategic decisions regarding player and team performance. Furthermore given the sports data rich environment and its increase in popularity over the past decade, cricket has recently seen an increase in analytical literature and the adoption of predictive methodologies at the professional level.

RESEARCH MOTIVATION

There is a scarcity in literature surrounding team rating systems utilising individual ability. This demonstrates a lack of demand and reflects a historical lack of access to data and computing resources. The research objective was to develop a roster-based optimisation system for T20 by deriving a meaningful, overall team rating using a combination of individual ratings from a playing eleven. The optimal team was defined as the set of 11 individual players that produce the greatest probability of winning for team, $i$, against any given opponent, $j$.

It was hypothesised that a team rating system accounting for individual player abilities, outperforms systems that only consider macro variables such as home advantage, opposition strength and past team performances. This research centres on the development of an adaptive-predictive rating system, characterised by utilising past player performances, and accounting for the long and short term variability of a team's performance. An adaptive method was preferred as it updates player and team ratings "based on historic performances upon availability of data about current performances" (Leitner, 2010, p.3). The assessment of system performance was observed through the prediction accuracy of future match outcomes, and benchmarked against the New Zealand Totalisator Agency Board (TAB) and CricHQ's predictive system (Bracewell *et. al.* 2014).

There are five key components in the development of the adaptive rating system: The first component was the data. The second component was the significant performance metrics. The third component was the optimisation system. The fourth component was the individual player rating system. The fifth component was the models ability to generate the probability of winning.

DATA

The analysis required end-of-match scorecard data for T20 cricket. Data was extracted from Cricinfo (www.espncricinfo.com). The developed system was tested on the Indian Premier League 2015. The scorecard data was split into a batting and bowling dataset outlining performance metrics, by player.

## 2. SIGNIFICANT PERFORMANCE METRICS

A random forest technique was introduced to handle multicollinearity and complex interactions to identify performance metrics that significantly affect a player's contribution to team winningness. Significant performance metrics were derived in terms of winningness (i.e. proportion of wins). The five most important metrics were: strike rate, balls faced, batting average, total runs scored and percentage boundaries. Percentage boundaries (batsmen) is defined as total boundaries divided by total balls faced. Interestingly these important metrics are associated with scoring efficiency (i.e. strike rate and percentage boundaries), scoring consistency (i.e. batting average) and scoring volume (i.e. total runs scored). The five most important bowling metrics were: economy rate, bowling average, strike rate, percentage boundaries and percentage dots. Interestingly, these important metrics are associated with wicket-taking efficiency (strike rate and bowling average), boundary prevention (i.e. percentage boundaries) and run restriction (i.e. economy rate and percentage dots). Percentage boundaries (bowlers) is defined as total boundaries conceded divided by total balls bowled, while percentage dots is defined as total dots divided by total balls bowled. The results show that reducing the number of runs conceded and increasing the rate at which wickets are taken are significant to winningness.

## 3. BINARY INTEGER PROGRAMMING

The optimisation method required the implementation of a binary decision variable, assigning a value = 1 to selected players and value = 0 otherwise (i.e. not selected). Since the adaptive rating system requires selecting players associated with the largest individual ratings, given a set of team and player-type constraints, a maximisation objective function is implemented. A Binary Integer Programming Model was adopted with the following framework:

The BIPM objective function:

$$Z = \sum_{i=1}^{n} \sum_{j=1}^{n_i} c_{ij} x_{ij} \,,$$

where $c_{ij}$ represents the player rating for player $j$ in role $i$, $\{i = 1,2,3,4\}$ ,

$$where\ role\ i = \begin{cases} 1, if\ batting\ ability \\ 2, if\ bowling\ ability \\ 3, if\ all - rounder\ ability \\ 4, if\ wicket\ keeping\ ability \end{cases}$$

Decision Variable:

$$x_{ij} = \begin{cases} 1, if\ player\ j\ is\ selected\ for\ role\ i \\ 0, otherwise \end{cases}$$

The decision variable are binary identifiers for player-type *i* and player *j*, where (*i=1,2,3,4)* and *(j=1,2,..,*$n_i$*).*

MODEL CONSTRAINTS

Model constraints that accurately reflect a team's composition and the type of talent required to win T20 cricket matches were assessed. The constraints must take into account the number of batsmen, bowlers, all-rounders, wicket-keepers and number of players required to build a cricket team. Given that model constraints were team orientated rather than individual player constraints, performance metrics that contribute significantly towards winningness at the team level, as opposed to the individual level, were established. The constraints were formulated such that the 'optimal' team produces the greatest probability of winning.

Applying the random forest technique the results indicated that batting metrics were of greater importance than bowling metrics for winningness among T20 teams. The results showed that seven of the top ten metrics were batting orientated, and predominately geared around scoring efficiency and consistency. It was revealed that batsmen with high scoring efficiency and scoring consistency are necessary to increase a team's chance of winning a T20 cricket match. Moreover, the results indicate that the model constraints should be formulated such that the optimal team generated by the optimisation system has a greater batting focus than bowling focus.

The constraints persuade the model to produce an optimal team with a heavy focus on batting ability as the model constraints require the optimal team to possess a greater number of batsmen than bowlers.

| **Constraints** | **T20** |
|---|---|
| *Team* | $\sum_{i=1}^{4} \sum_{j=1}^{n_1} x_{ij} = 11$ |
| *Player* | $\sum_{i=1}^{4} x_{ij} \leq 1$ |
| *Batsmen* | $\sum_{j} (x_{1j} + x_{3j}) \geq 7$ |
| *Bowler* | $\sum_{j} (x_{2j} + x_{3j}) \geq 4$ |
| *All-rounder* | $\sum_{j} x_{3j} \geq 0$ |
| *Wicket-Keepers* | $\sum_{j} x_{4j} = 1$ |

Table 1: Player-type and Model Constraints

## 4. EVALUATING INDIVIDUAL PLAYER AND TEAM RATINGS

The individual rating method implemented into the adaptive rating system was a combination of the Product Weighted Measure + Analytical Hierarchy Process and Exponentially Weighted Moving Averages.

The optimal team rating was calculated by aggregating individual player ratings of the selected players. This aggregation approach was justified in (Damodaran, 2006), stating that cricket is a sport characterised by one-on-one interactions between batsmen and bowlers, and that a players ability establishes the outcome of this interaction. Moreover the match outcome is defined by the interactions between batsmen and bowlers, therefore summing the individual player ratings provides a fair indication of team strength. Once ratings for team $i$ and $j$ have been calculated, the Bradley-Terry model was applied to calculate the probability of team $i$ beating team $j$:

$$\pi_{i,j} = \frac{R_i}{R_i + R_j}$$

Leitner (2010) stated that the outcome of many sporting disciplines can be determined by pairwise comparisons, and that the outcome of a match or game is dependent on the current ability of the two teams.

PRODUCT WEIGHTED MEASURE

The Product Weighted Measure (PWM) was developed and applied by Croucher (2000) to rank batsmen, bowlers, wicket-keepers and all-rounders in international one day cricket. The method produces raw ratings for each player and then calculates the actual ratings relative to other players within their player-type (please refer to [6] for further details on how each player-type rating is derived). However the performance metrics used to rank the players were selected in an ad hoc manner, and the weightings, were subjectively chosen. Given the difference in importance of each performance metrics, the author introduced a novel method, utilising the Analytical Hierarchy Process and Random Forest technique, to determine the appropriate weightings, α, for each important performance metric, for each player-type.

ANALYTICAL HIERARCHY PROCESS

The Analytical Hierarchy Process (AHP) is a multi-criteria decision making tool developed by Thomas Saaty (1987). Given a user defined pairwise comparison matrix, the AHP translates the matrix into a vector of relative weights for each criterion element using a mathematical model. The pairwise comparison matrix provides a numerical comparison of each attribute with respect to the other attributes being evaluated. These matrix entries are determined using the fundamental AHP scale and are based on prior experience or expert knowledge. Applying the AHP to the pairwise comparison matrix translate the subjective weights into objective weights, representing the importance of the attribute relative to the other attributes. Moreover the method implements a consistency measure for each attribute to ensure that the 'user' defined weights are consistent and reduces bias in the decision making process.

RANDOM FOREST + AHP WEIGHTINGS

The system for determining the appropriate weightings is outlined as follows:

1. Identify the order of importance for each performance metric, for each player-type.
2. Use the order of importance to create a $n \times n$ pairwise comparison matrix, for each player-type, where each entry, $a_{ij}$ represents the importance of criteria $i$ with respect to $j$. The relative importance of each performance metric, $a_{ij}$, follows the importance order established by the random forest importance plot. For example, if percentage boundaries are of greater importance to winningness than batting average, for batsmen, the relative importance of percentage boundaries versus batting average > 1.
3. Run the AHP on the pairwise comparison matrix and generate the weights associated with each performance metric for each player-type. The following weightings were generated:

| **Performance Metrics** | **Batsmen** | **Bowlers** | **All-rounders** | **Wicket-Keepers** |
|---|---|---|---|---|
| Total Runs Scored | 0.33 | - | 0.34 | 0.33 |
| % Boundaries (batting) | 0.30 | - | 0.30 | 0.30 |
| Batting Strike Rate | 0.37 | - | 0.36 | 0.37 |
| % Boundary (bowling) | - | 0.30 | 0.35 | - |
| Bowling Strike Rate | - | 0.33 | 0.27 | - |
| Economy Rate | - | 0.37 | 0.38 | - |

Table 2: Weightings for T20 cricket performance metrics

These weights align with findings established above, stating that a winning T20 team requires players with high scoring efficiency, high scoring consistency and high run restricting ability. The table shows that metrics such as batting strike rate, total runs scored and economy have a greater weighting relative to other metrics.

FORECASTING METHOD

Since the PWM ratings are generated relative to the sum of the other ratings, for a given player-type, this enables the ability to track player performance on a match-by-match basis, and assesses a player's progression as the season matures. The time-stamped ratings enabled the application of forecasting methods to player ratings. Daniyal *et al.* (2012) applied Exponentially Weighted Moving Average (EWMA) control charts to individual batting performances. Moreover, exponential smoothing was applied by Clarke (2011) to predict tennis player ratings. It was found that exponential smoothing produced predictive player ratings. Bracewell and Ruggiero (2009) utilised control charts to monitor batting performances of New Zealand domestic cricketers, and established that control charts such as EWMA accurately forecasted a batsmen's form.

SYSTEM ACCURACY

The EWMA methodology was embedded into the PWM individual rating method with a weighting measure of 0.72. This method predicts a players rating for the following match, and filters the predicted ratings through the optimisation system to generate a forecasted team rating. Applying this method to the Indian Premier League (2015) the following predictive accuracy was established:

| Competition | TAB | CricHQ | Adaptive System |
|---|---|---|---|
| Indian Premier League | 60% | 64% | 72% |

Table: 3: Predictive Systems Accuracy for 2015 IPL Tournament

**5. DISCUSSION AND CONCLUSION**

Given cricket's exponential growth into a multi-billion dollar industry, it has become more critical than ever to introduce analytical methods for team selection. The adaptive system is useful for decision making among coaching and managerial staff, in terms of player selection, and can be implemented to identify the optimal team for T20 cricket.

The lack of academic literature surrounding team rating systems utilising individual ability within cricket, the absence of the application of predictive techniques to forecast match outcome and the growing popularity of sports betting, established an entry point in the market for this research.

This research developed a roster-based optimisation system for T20 cricket by deriving a meaningful, overall team rating using a combination of individual ratings from a playing eleven. The research revealed that an adaptive rating system accounting for individual player abilities, outperforms systems that only consider macro variables such as home advantage, opposition strength and past team performances. The assessment of system performance was observed through the prediction accuracy of future match outcomes.

The adaptive rating system was applied to the Indian Premier League 2015, and the systems predictive accuracy was benchmarked against the New Zealand Totalisator Board Agency (TAB) and the CricHQ algorithm.

The results revealed that the developed rating system outperformed the TAB and CricHQ algorithm by 20% and 13%, respectively. The result demonstrates that cricket team ratings based on the aggregation of individual player ratings are superior to ratings based on summaries of team performances and match outcomes; validating the research hypothesis. This demonstrated that rating systems that consider micro variables generate greater predictive accuracy than systems that only consider macro variables.

**References**


Annis, D. H., and Craig, B. A. (2005). *Hybrid paired comparison analysis, with applications to the ranking of college football teams*. Journal of Quantitative Analysis in Sports, 1(1).

Ghosh, P. (2014). Billions of dollars at stake: Why is the international cricket council changing its revenue sharing model? http://www.ibtimes.com/billions-dollars-stake-why-international-cricket-council-changing-its-revenue-sharing-1554078. Accessed: 2016-02-05.

Bracewell, P. J., Downs, M. C. F., & Sewell, J. W. (2014). *The development of a performance based rating system for limited overs cricket*. In MathSport 2014, 40-47.

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

Clarke, S. R. (2011). *Rating non-elite tennis players using team doubles competition results*. Journal of the Operational Research Society 62(7), 1385–1390.

Fenez. M, & Clark, J.D. (2011). *PWC outlook for the global sports market to 2015*. Report P-25, PriceWaterhouseCoopers.

Croucher, J. (2000). *Player ratings in one-day cricket*. In Proceedings of the fifth Australian conference on mathematics and computers in sport. Sydney University of Technology, Sydney, NSW, 95–106.

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

Daniyal, M., Nawaz, T., Mubeen, I., and Aleem, M. (2012). *Analysis of batting performance in cricket using individual and moving range (MR) control charts*. International Journal of Sports Science and Engineering 6(4), 195–202.

Di Salvo, V., Baron, R., Gonzalez´-Haro, C., Gormasz, C., Pigozzi, F., and Bachl, N. (2010). *Sprinting analysis of elite soccer players during European Champions league and UEFA cup matches*. Journal of Sports Sciences 28(14), 1489–1494.

Harville, D. A. (2003). *The selection or seeding of college basketball or football teams for postseason competition*. Journal of the American Statistical Association 98(461), 17–27.

Gerber, H., & Sharp, G. (2006). *Selecting a limited overs cricket squad using an integer programming model*. Journal of the Operations Research Society, 5.

Leitner, C., Zeileis, A., & Hornik, K. (2010). *Forecasting sports tournaments by ratings of (prob) abilities*: A comparison for the euro 2008. International Journal of Forecasting, 26(3), 471–481.

Manage, A. B., & Scariano, S. M. (2013). *An introductory application of principal components to cricket data*. Journal of Statistics Education 21, 3.

Saaty, T. (1987). *The analytic hierarchy process*. Math Modelling, 9, 161–176.

Sargent, J. (2013). *Player ratings in continuous and discrete team sports*. PhD thesis, RMIT University.

Schumaker, R. P., Solieman, O. K., & Chen, H (2010). *Predictive modeling for sports and gaming*. Springer, 2010.

Sharp, G., Brettenny, W., Gonsalves, J., Lourens, M., & Stretch, R. (2011). *Integer optimisation for the selection of a Twenty20 cricket team*. Journal of the Operational Research Society, 62(9), 1688–1694.

Sinuany-Stern, Z. (1988). *Ranking of sports teams via the AHP*. Journal of the Operational Research Society, 661–667.

Stefani, R. (2011). *The methodology of officially recognized international sports rating systems.* Journal of Quantitative Analysis in Sports, **7**(4).

Vrooman, J. (2012). The economic structure of the NFL. In *The Economics of the National Football League*. Springer, 7 31.

# THE DUCKWORTH-LEWIS-STERN METHOD: RAIN RULES AND MODERN SCORING IN ONE-DAY CRICKET

Steven E. Stern [a,b]

[a] *Queensland University of Technology (QUT)*
[b] *Corresponding author: steven.stern@qut.edu.au*

## Abstract

The famous (and occasionally infamous) Duckworth-Lewis methodology for dealing with interruptions in limited overs cricket matches made its international debut in early 1997. For 20 years, it has set the standard for target adjustment at nearly all levels of the game. In that time, it has not been static. In 2003, the Professional Edition of the method was introduced to handle changes to scoring patterns which were becoming apparent in modern cricket. We here detail the Duckworth-Lewis-Stern (DLS) method, adopted in 2014 and designed to deal with the now common extreme scoring rates seen in limited overs matches, particularly Twenty20. In addition, we outline key principles governing the structure and properties of target adjustment methods and compare DLS to other proposed procedures.

*Keywords*: Cricket; Scoring Resources; Sports modelling

### Acknowledgements

We wish to thank Frank Duckworth and Tony Lewis for their ongoing support as well as the ICC.

### References

Bhattacharya, R; Gill, PS & Swartz, TB (2011), Duckworth-Lewis and Twenty20 cricket. The Journal of the Operational Research Society 62, 1951-1957.

Brooker, S & Hogan, S (2012). "Cricket and the Wasp: Shameless self promotion (Wonkish)". http://offsettingbehaviour.blogspot.co.nz/2012/11/cricket-and-wasp-shameless-self.html

Clarke, SR & Allsopp, P (2001). Fair measures of performance: The World Cup of cricket. The Journal of the Operational Research Society 52, 471-479.

Duckworth, FC & Lewis, AJ (1998). A fair method of resetting the target in interrupted one-day cricket matches. The Journal of the Operational Research Society, 49, 220-227.

Duckworth, FC & Lewis, AJ (2004). A successful operational research intervention in one-day cricket. The Journal of the Operational Research Society 55, 749-759.

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

Jayadevan, V (2014). VJD method software. http://jayadevan.yolasite.com/cricket.php.

McHale, IG & Asif, M (2013). A modified Duckworth-Lewis method for adjusting targets in interrupted limited overs cricket. European Journal of Operational Research 225, 353-362.

Perera, HP & Swartz, TB (2012), Resource estimation in T20 cricket. IMA Journal of Management Mathematics 24, 337-347.

Preston, I & Thomas, J (2002). Rain rules for limited overs cricket and probabilities of victory. The Statistician (Journal of the Royal Statistical Society: Series D) 51,189-202.

Schall, R & Weatherall, D (2013). Accuracy and fairness of rain rules for interrupted one-day cricket matches. Journal of Applied Statistics 40, 2462-2479.

# MODELLING THE THIRD VERSION OF THE CYCLING OMNIUM

Loic Devoir[a], Bahadorreza Ofoghi[b], Dan Dwyer[c], Clare MacMahon[d], and John Zeleznikow[a,e,f]

[a] *Institute of Sport, Exercise, and Active Living, Victoria University, Australia*
[b] *Department of Computing and Information Systems, The University of Melbourne, Australia*
[c] *Deakin University, Australia*
[d] *Swinburne University, Australia*
[e] *College of Business, Victoria University, Australia*
[f] *Corresponding author: john.zeleznikow@vu.edu.au*

**Abstract**
This paper reports our recent findings on the analysis of riders' performance patterns in the new (third version) Track Cycling Omnium. We have made use of statistical and machine learning techniques to analyse data in recent omnium competitions and compared the results with our previous findings on the older version of the omnium to understand how and whether the new omnium requires different skill sets and strategic planning for elite riders and their coaches. The results of our analysis shows that for both male and female riders, sprint abilities are more important than endurance abilities in the current omnium, contrary to the previous omnium where endurance abilities were slightly more prominent in female medal winners. In addition, the Flying Time Trial race remains the most important race for men while Individual Pursuit has become more important for female riders.

**Keywords**: Track cycling omnium, rule changes, performance analysis

## 1. INTRODUCTION

This work focuses upon track cycling omnium coaches as decision makers who identify talent and conduct strategic performance planning for cyclists that compete in the multi-event track cycling omnium. Such advice is given to coaches/riders before and during the omnium events. The goal is to select the best performing cyclists and support them to finish in the highest possible position in the competition.

Track cycling omnium is a competition that was introduced by the International Cycling Union (UCI) in 2007. It originally consisted of five individual events. In December 2009, the UCI announced new changes to the omnium competition (addition of an extra event, the elimination race, and held over two days rather than the previous one) that took place for the first time in the 2010–2011 track cycling season and the 2011 World Cycling Championships. It was contested for the first time at an Olympic Games in London in 2012. The individual events in the second version of the omnium included:

- Flying Time Trial: 250m for men and women: Cyclists compete in a short flying lap that has traditionally been used for qualification in other cycling events (e.g., sprint competitions). Each rider completes their FTT separately.
- Points Race: 30km for men and 20km for women: This is a mass start event involving a large number of riders on the track at the same time. Every 10 laps, a sprint is held and the top four finishers are awarded 5, 3, 2, and 1 point(s), respectively. Any rider who takes a lap on the field is awarded 20 points for each lap taken. The rider with the most points is the winner of the event.
- Elimination Race: 24 riders and 50 laps for men and women: 24 riders compete in a race where every two laps, the last rider over the finish line is eliminated until only a single rider remains and is decided as the winner.
- Individual Pursuit: 4km for men and 3km for women: Two riders start a race from opposite sides of the track on the pursuit line at the bottom of the track. The riders start at the same time and both must complete the race distance to record a time for the ranking relative to the other cyclists.
- Scratch Race: 15km for men and 10km for women: All contestants start from a start point at the same time and need to complete a certain number of laps. If a rider gains an entire lap ahead of the other riders, she/he will have an advantage over the riders who have completed fewer laps.
- Kilometer Time Trial: 1km for men and 500m for women: Riders compete against the clock to secure the fastest time. Riders are sent out individually in set time intervals. Whilst still considered a sprint, the distances are considerably longer than those in the flying time trial event.

Each omnium has 24 competitors. The winner of each event receives, 1 point, second 2 points etc. The points for each of the 6 events are summed together. The winner of this version of the cycling omnium was the competitor with the least number of points.

The mixture of skills and levels of performance that is required in multiple-component competitions makes it difficult to determine what strategies and/or athletes best suit a plan to maximize the possibilities of winning medals in such competitions (Ofoghi, Zeleznikow, MacMahon, & Dwyer, 2010). Some example studies on multiple-component sports include the works by Zwols and Sierksma (2009), Kenny, Sprevak, Sharp, and Boreham (2005), and Cox and Dunn (2002) on decathlon, as well as Ofoghi, Dwyer, Zeleznikow, MacMahon, and Rehula (2016) and Cejuela, Pérez, Villa, Cortell, and Rodríguez (2008) on triathlon.

In respect of track cycling omnium, we carried out a number of analyses in both of the old five-event and the second version (six-event) omniums to find winning performance patterns in the omnium with the aim of assisting coaches in decision making towards selecting the most appropriate riders to compete in the contest (Ofoghi, Zeleznikow, Dwyer, & MacMahon, 2013; Ofoghi, Zeleznikow, MacMahon, & Dwyer, 2011; Ofoghi et al., 2010), and to find the optimal performance patterns in terms of rankings in the omnium individual events before and during the omnium (Ofoghi, Zeleznikow, MacMahon, & Dwyer, 2013) to help cyclists strategize for finishing in the best overall position at the completion, given their performances to specific stages in the omnium. We also analysed the individual event Elimination Race to better understand how riders can avoid elimination in different stages of the event (Dwyer, Ofoghi, Huntsman, Rossitto, MacMahon, & Zeleznikow, 2013).

In 2014, the omnium went through a third round of changes made by the UCI. Effective from June that year, the current omnium includes the same six events as in the second version of the omnium, held over 2 days. However, the order of the events has changed and it now consists of the Scratch Race, an Individual Pursuit, an Elimination Race, a Kilometer Time Trial, a Flying Time Trial, and a Points Race. For the first five events, the winner is awarded 40 points, the rider in the second place receives 38 points, etc. Riders who are ranked 21$^{st}$ and below are awarded 1 point. In the sixth event (Points Race), cyclists add to and lose points from their points total based on the laps gained and lost (plus and minus 20 points) as well as the points they win in the sprints of the race (A sprint is held every ten laps, with 5, 3, 2, and 1 point(s) being awarded to the top four finishers in each sprint). The winner of the omnium is the rider who obtains the highest total points. To win the omnium, a rider must have completed every individual event in the omnium.

The current research aimed at building on previous work on supporting decision making in the older versions of track cycling omnium and further developing the understanding of performance patterns in the current omnium in effect as of June 2014. More specifically, this paper has a focus on finding winning patterns in the current omnium, for both female and male riders, using statistical and machine learning techniques.

From a practical point, we are interested in whether the rule changes for the third version of the cycling omnium have led to changed outcomes.

## 2. METHODS

### RESEARCH QUESTIONS

The first step of our analysis was to determine how the new rules effective since 2014 have impacted the omnium competition. This led to the question of how the riders would have finished in previous omniums if they had to compete with the rules that apply to the current version of the omnium. Second, our intuition was that the overall score that riders collect in the Points Race has now become more important than before as this is currently the last individual event in the omnium. Therefore, another question was to determine if riders reaching a high score in the Points Race will be more influential, than previously, in the overall standings. As a generalization of this question, we then decided to find the significance of each individual event for the overall standings of riders in the new omnium as compared with the previous omnium. Following our previous work on the second version of the omnium, we also wanted to understand which type of cycling expertise, i.e., sprint or endurance abilities, plays a more important role in the current omnium for both male and female riders.

### DATA COLLECTION

To find answers to the afore-mentioned questions, we collected the results of all omnium competitions since 2010 in the UCI World Cup, UCI World Championship events and the 2012 Summer Olympics. We decided to not consider the results of competitions prior to 2010 as there is a significant gap between the scoring systems of the first and the third versions of the omnium. Further, the second and third versions of the omnium include the same six events and are held over two days. Therefore, we only focused on the results of the second and third versions of the omnium for all of the analyses in this paper. We only looked at the rankings of riders in each individual event and their overall placing for the first version of the omnium. For the current (third) version of the omnium, we collected the scores of cyclists in the individual events and overall.

## 3. ANALYSIS AND RESULTS

To answer the first research question on the impact of the new rules on the omnium, we converted the results of all the previous omniums (i.e., the second version of the omnium) to the scores the riders would have collected with the new scoring system in the omnium. For this, the rankings of riders were converted to the scores they would have achieved in each individual event. The overall standings of riders were then inferred by comparing the overall scores of the riders calculated using this new scoring system. Then, statistical measures were used to determine whether the differences between the overall standings of riders with the old and new scoring systems are significant. The results of this analysis are shown in Table 1.

| Statistical method | Men | Women |
|---|---|---|
| KL Divergence | 91.4236 | 35.1771 |
| Paired T-Test (*p*-value) | 0.9050 | 1.0 |
| T-Test (*p*-value) | 0.9823 | 1.0 |

**Table 1:** The results of statistical analyses of the differences between the overall rankings of riders in previous omniums with the old and current scoring systems

The results shown in Table 1 show that the differences between the overall rankings of riders with the scoring system that was in place at the time compared with the rankings that the riders would have achieved with the current scoring system are not statistically significant. This suggests that the scoring system would have made no significant impact on the overall standings of neither male nor female riders in the omnium even if the new scoring system had to be applied to the rankings and scores of the riders.

The second round of the analysis was focused on the significance of the Points Race specifically and all the other individual events in the omnium. The Pearson Correlation measure was used in this case to measure and compare the correlation of the results of riders in each of the individual events with the overall placings of riders.

| Individual event | Men | | Women | |
|---|---|---|---|---|
| | omnium II | omnium III | omnium II | omnium III |
| Flying Time Trial | 0.84 | 0.81 | 0.70 | 0.65 |
| Points Race | 0.68 | 0.47 | 0.58 | 0.65 |
| Elimination Race | 0.64 | 0.68 | 0.66 | 0.63 |
| Individual Pursuit | 0.74 | 0.79 | 0.80 | 0.77 |
| Scratch Race | 0.63 | 0.67 | 0.66 | 0.64 |
| Kilometer Time Trial | 0.70 | 0.73 | 0.82 | 0.76 |

**Table 2:** The results of the correlation analysis of the performance of riders in each individual event with their final standings in the second and third versions of the omnium, i.e., omnium II and omnium III

The correlation results in Table 2 suggest that the new omnium rules as applied in the current version of the omnium has had a mixed impact on the significance of the performances of riders in the Points Race on the overall standings of the riders. For men, the Points Race is now less correlated with the overall rankings of riders whereas for female riders, it has become more associated with the final standings in the omnium.

In addition, the results in Table 2 demonstrate that for men the new rules have had no significant impact on the other five individual events. This can be observed by comparing the correlation measures for omnium II and omnium III. Nevertheless, the Individual Pursuit and Kilometer Time Trial events have become slightly more important. And among all of the individual events, the Flying Time Trial is still the most correlated individual event with the overall omnium rankings, contrary to the intuition that the Point Race would be the most correlated event with the final standings.

For female riders, the correlations of the rankings in the other five individual events with the overall placings have changed almost to the same extent as for men. The Individual Pursuit has now become the most correlated race with the final standings contrary to omnium II in which the 500 Metre Time Trial was the most important event. And similar to what was found in the results of male riders, in women’s omnium III competitions, the Points Race is not the most important race of the six in the omnium either.

To understand which type of cycling expertise is more important for male and female riders to finish the current omnium with a medal, we divided the riders in three categories: i) riders on the podium, ii) riders ranked between 4 and 10, and iii) those who finished in places above 10. Then, we adopted the same methodology as in our previous work (Ofoghi et al., 2011) to calculate the *mean sprint score* (mss) and *mean endurance score* (mes). For the current analysis, however, we have changed the formulae of the mss and mes scores as shown in Equation 1 and Equation 2, respectively, where $E_{mscore}$ is the mean of the scores of riders in the event $E$ which was calculated using the K-means clustering algorithm (MacQueen, 1967). The change in the calculation of the mss and mes scores was based on advice from a track cycling expert.

$$mss = \frac{FTT_{mscore} + KTT_{mscore}}{2} \tag{1}$$

$$mes = \frac{SR_{mscore} + IP_{mscore}}{2} \tag{2}$$

To make the mss and mes scores comparable across the two versions of the omnium, we transformed the rankings of riders in the second version of the omnium to scores (i.e., rank 1 replaced with 40 points , rank 2 replaced with 38 points, etc. as explained in the Introduction).

The results of the mss and mes score analysis are summarized in Table 3. From these results, it can be observed that sprint ability is still more important for male riders who finish the current omnium with a medal (as in the previous omniums). This is understood when comparing the mss scores and mes scores for the riders in the same omnium version III. The difference between the mss and mes scores in the new omnium is approximately 2 points (33.18 vs. 31.32) which eventuates to 1 rank difference. From the second version of the omnium to the current version, however, the overall sprint ability required for winning a medal has slightly decreased (from 33.53 to 33.18).

| Gender/version | mss | | | mes | | |
|---|---|---|---|---|---|---|
| | Podium | fs 4-10 | fs >10 | podium | fs 4-10 | fs >10 |
| male/II | 33.53 | 30.03 | 12.78 | 31.81 | 25.96 | 14.01 |
| male/III | 33.18 | 24.86 | 12.93 | 31.32 | 26.26 | 12.85 |
| female/II | 32.08 | 26.35 | 12.44 | 32.66 | 25.13 | 13.64 |
| female/III | 34.74 | 26.10 | 13.19 | 33.43 | 25.72 | 14.19 |

**Table 3:** The sprint vs. endurance ability analysis of omnium II and omnium III competition results for men and women. Note: fs = final standing, mss = mean sprint score, and mes = mean endurance score.

More importantly, from Table 3, for female riders, the endurance ability is not the more important expertise required to finish the current omnium with a medal anymore. In the second version of the omnium, the endurance ability was slightly more important for medal winners (32.08 in omnium II vs. 32.66 in omnium III); however, in the current omnium, sprint ability has been shown to be slightly more important as indicated by the difference between a mss of 34.74 vs. a mes of 33.43 for riders on the podium.

## 4. DISCUSSION

As with our previous research on the cycling omnium, our research does not concur with the intuition of cycling experts. For example, the six-event Omnium was supposedly intended to give endurance riders a chance to compete and win medals in track cycling According to Cycling Weekly "*The format for the Olympic Games Omnium event has been confirmed by the UCI, with the elimination race being added to make the competition more conducive to endurance riders*"; (Birnie, 2009). In Ofoghi et al. (2013a), we noted that although most statistical test results are not significant, the addition of the elimination race event to the second version of the omnium seems to have failed to achieve the goal of bringing more opportunity of medal winning to endurance riders.

Anderson (2014) notes that "*the riders' overall points tally will be taken in to the final round points race and added to, or subtracted from (should they lose a lap), meaning spectators can follow the overall classification as it happens*". According to Guardian (2014), "*the points race is even more important now after the UCI, cycling's world governing body, reshuffled the order of the omnium disciplines*". Similarly, our intuition was that the overall score that riders collect in the Points Race should now become more important than before as the Points Race is currently the last individual event in the omnium. But surprisingly, riders reaching a high score in the Points Race did not achieve (compared to previous outcomes), a higher score in the overall standings. Indeed, we found that differences between with the current and previous scoring systems are not statistically significant and that the changed scoring system has no discernible impact on the overall standings of either male or female riders in the omnium

On the other hand, the new rules do not seem to have changed the dynamics and the importance of each of the individual events of the omnium in the overall success of male riders, as evident in the correlation measures summarised in Table 2; whereas for female riders this has resulted in some minor shift towards the Individual Pursuit.

When focusing on the two types of cycling expertise, i.e., sprint vs. endurance, overall, sprint ability has been shown to be slightly more important in the current omnium for both genders, contrary to the previous version of the omnium where medal winning female riders were required to have more endurance power. This is a slight paradigm shift for the coaches of female omnium competitors as well as the riders who target podium places in the current omnium events.

## 5. CONCLUSIONS

We carried out statistical and machine learning analysis on the competition results in the third version of the track cycling omnium and compared the results with those on the second version of the omnium. The results of our analyses show that the current omnium has had some impacts on the performance patterns of both male and female riders, especially those who finish on the podium. The most important individual events are now the Flying Time Trial and Individual Pursuit for men and women, respectively. Any contrary to our intuition, the performances of riders in the final individual event, i.e., the Points Race, do not seem to have the highest impact on winning the omnium. On the other hand, while the sprint ability was only more important for men in the previous version of the omnium to win a medal, in the current omnium, both men and women are required to have slightly higher sprint abilities than endurance power in order to secure a greater chance of finishing the omnium in a medal winning place.

## REFERENCES


Anderson, G. (2014). UCI announces change to points system in omnium and team pursuit disciplines, http://www.insidethegames.biz/articles/1021015/uci-announces-changes-to-points-system-in-omnium-and-team-pursuit-disciplines, last accessed: May 9 2016.

Birnie, L. (2009, December 17). Cycling Weekly. Retrieved from http://www.cyclingweekly.co.uk/news/latest/435675/olympic- Omnium-format-unveiled.html, last accessed: 9 May 2016.

Guardian (2014). Laura Trott relishing omnium changes and new focus on points race, http://www.theguardian.com/sport/2014/oct/21/laura-trott-points-race-omnium-track-cycling, last accessed: 9 May 2016.

Kenny, I.C., Sprevak, D., Sharp, C., & Boreham, C. (2005). Determinants of success in the Olympic decathlon: Some statistical evidence. *Journal of Quantitative Analysis in Sports*, 1.

Cox, T.F., & Dunn, R.T. (2002). An analysis of decathlon data. *Journal of the Royal Statistical Society*, 51, 179–187.

Zwols, Y., & Sierksma, G. (2009). Training optimization for the decathlon. *Journal of Operations Research*, 57, 812–822.

Bahadorreza Ofoghi, Dan Dwyer, John Zeleznikow, Clare MacMahon, & Jan Rehula (2016). Performance analysis and prediction in triathlon. *Journal of Sports Sciences*, 34(7), 607-612.

Cejuela Anta, R., Pérez Turpin, J., Villa Vicente, J., Cortell Tormo, J., & Rodríguez Marroyo, J. (2008). An analysis of performance factors in sprint distance triathlon. *Journal of Human Sport and Exercise*, 2(2), 1-25.

Dwyer, Dan, Ofoghi, Bahadorreza, Huntsman, Emy, Rossitto, Daniel, MacMahon, Clare, & Zeleznikow, John (2013). The elimination race in track cycling: Patterns and predictors of performance. *Journal of Science & Cycling*, 2(2), 6-12.

Ofoghi, Bahadorreza, Zeleznikow, John, MacMahon, Clare, & Dwyer, Dan (2013a). Supporting athlete selection and strategic planning in track cycling omnium: A statistical and machine learning approach. *Information Sciences*, 233, 200-213.

Ofoghi, Bahadorreza, Zeleznikow, John, Dwyer, Dan, & MacMahon, Clare (2013b). Modeling and analysing track cycling omnium performances using statistical and machine learning techniques. *Journal of Sports Sciences*, 31(9), 954-962.

Ofoghi, Bahadorreza, Zeleznikow, John, MacMahon, Clare, & Dwyer, Dan (2011). Has the addition of the elimination race to the track cycling omnium benefitted sprinters or endurance riders? In *Proceedings of the International Symposium on Computer Science in Sport*, Shanghai, China.

Ofoghi, Bahadorreza, Zeleznikow, John, MacMahon, Clare, & Dwyer, Dan (2010). A machine learning approach to predicting winning patterns in track cycling omnium. In *Proceedings of the International Federation for Information Processing (IFIP) Conference on Advances in Information and Communication Technology*, pp. 67-76, Brisbane, Australia, 2010.

MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations. In *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press. pp. 281–297.

# THE INTERRELATED EVOLUTION OF SEVEN FOOTBALL CODES: THE ROUGE, THE RUBBER BLADDER, WEBB ELLIS AND TOM WILLS

Ray Stefani [a,c] , Richard Pollard [b]
[a] *California State University, Long Beach, USs*
[b] *California Polytechnic State University, San Luis Obispo, USA*
[c] *Corresponding author: Raystefani@aol.com*

## Abstract

Today's diverse football codes share similarities (kicking goals, the fair catch, offside and two types of an early rule called the rouge) by which a fascinating evolution can be traced much as the migration of humankind can be traced by DNA. Prior to 1800, football in England was a rough, continuous-action sport with many variations. Only goals counted. After 1800 two evolutionary paths emerged. In the "schools" path, Eton and others championed what become the Football Association or FA (1866) kicking game while Rugby and others led to the creation of the running game and to the formation of the Rugby Union (1871). Playing conditions and the shape of the available ball were important stimuli. Both codes allowed a fair catch and the Eton rouge until 1866. Rugby reached Canada in 1861. The American code combined versions of the kicking game from 1869-1875 and then switched to the Canadian rugby code in 1876. Scoring rules were adopted by the American and Canadian codes in 1883 and by Rugby Union in 1886. Professional rugby (Rugby League) began in England in 1895, the same year that professional football began in the American code. Tackling/heeling entered American (1882), Canadian (1887) and Rugby League (1906) codes. The forward pass was legalized by American (1906) and Canadian (1929) codes. The "public" path led to Australian Rules (1859) and Gaelic (1888) football, both purposely different from other codes with the Sheffield rouge and no offside rule. Today's codes share a remarkably interactive history..

**Keywords**: Soccer, Rugby Union, Rugby League, American football, Canadian football, Australian Rules football, Gaelic football

### 1. INTRODUCTION

A casual examination of today's major football codes reveals obvious similarities among pairs of codes such as with American and Canadian football codes, with Rugby Union and Rugby League and with Australian and Gaelic football codes. On the other hand, Association football is clearly different from those six codes in that handling is only allowed by the goalkeeper. How did those codes arrive at their current forms? A similar question can be asked about the races of humankind. Obvious physical regional similarities exist, but how did those races reach their current locations? As to humankind, DNA is used. Common sets of DNA indicate common ancestry while different DNA sequences indicate which races separated in the march "out of Africa". Further, tool making is well known to have accelerated and affected the growth and distribution of humankind.

Rules are the DNA of football. I believe that rules such as the Eton Rouge and Sheffield Rouge have been overlooked as markers of the evolution of football codes. Those rules will be discussed as one contribution of this paper. As football evolved "out of England" the tool was of course the football, the shape of which I believe is overlooked as a prime cause of the early evolution of codes, another contribution herein.

This paper begins with the separation of football into kicking and running codes in England with a focus on the root causes, not just on dates. The story of William Webb Ellis as a possible influence of rugby will be briefly discussed. The evolutionary story then moves to American and Canadian codes where scoring rules demonstrate interactivity of code evolution. Rugby League is covered with other rules evolutions. Finally Australian and Gaelic codes appear to have purposely followed Sheffield and not London/FA rules. A concluding figure shows a side-by-side comparison of codes much as has been done using DNA for humankind. Due to space limitations, a few of almost 100 references are placed at the end.

### 2. RUNNING AND KICKING CODES EVOLVE AFTER 1800

Prior to 1800, there appear to be two versions of football. An earlier London dispatch quoted in the 2 April 1733 edition of *The Weekly Rehearsal*, a Boston newspaper, describes a football playing mob invading

Christmas Day services in Chester by kicking a ball down the main isle and putting out candles. On the other hand, a 14 February 1736 article in the *New York Weekly Journal* states "Two to one is odds at Foot-ball". That article indicates that odds such as two to one had already entered the vernacular and that football had one version that was organized well enough so that a prediction of the outcome was possible. Few rules were written down; but, generally the ball was pushed or kicked forward until it was propelled through a goal.

FACTORS INFLUENCING THE RULES

By 1800, football had evolved in two directions: schools and clubs. The English public school system began to use that game to teach sportsmanship; however, the form of play had to be safe for England's future leaders. I believe there were three main influences on the evolution of the school game: ego, the type of playing field and the available ball. On one hand, Eton, Winchester, Harrow and others played in cloisters or in confined areas. These schools favoured more of a kicking/dribbling game. Conversely, Rugby, Marlborough, Cheltenham and others had larger grassy playing areas that supported running with the ball and tackling. Rugby School had (and still has) the famous "Close". A close was the $17^{th}$ and $18^{th}$ century term for a large playing field enclosed by some marked boundary. The ego (or leadership) of Eton and Rugby mandated that their games were to remain different, one emphasizing kicking and the other running, mainly due to safety considerations; however, a number of rules were common to both codes.

The third influence was the ball. Prior to about 1820, the ball was composed of a pig bladder, blown up by lung power. By 1823, William Gilbert of Rugby, England produced a ball with a pig bladder covered by a leather exterior. The current Gilbert Museum has an exhibit with a pig-bladder ball and an early Gilbert ball. The discussion by a contemporary player displayed by the Museum indicates that the resulting plum-shaped ball could either be dribbled or handled with equal ease. I believe this multi-purpose ball allowed the Eton and Rugby factions to employ similar rules. There were four common rules to the kicking and running codes from about 1820 to about 1866.

- Goals were scored by kicking
- Fair catch/mark
- Offside was determined by the location of the ball
- Eton Rouge

The fair catch involved catching the ball in mid air, making a mark where the ball was caught and making a free kick for goal from that mark. The Eton Rouge survives in today's codes (with varying outcomes) and is much overlooked as a common feature. Eton Rules were first published in 1815. If the ball was kicked into the in-goal area and the attacking team touched it first, the attacking team was awarded a free kick at goal. If the defending team touched the ball first, the defending team was awarded a free kick away from its goal. Kicking the ball into the in-goal area to create a score is an obvious feature of Rugby Union today, but was then also a feature of the kicking game that eventually became Association football.

RULES PROLIFERATE AND CODES SEPARATE

Rules were published starting in 1815 by proponents of the kicking game and by proponents of the running game. One source of rules was the Sheffield Football Club whose rules evolved from the clubs path rather than the schools path

- Kicking Code
  - Eton 1815, 1847
  - Cambridge 1856
  - Uppingham 1862
  - Sheffield 1858, 1861, 1863, 1867, 1870
  - Football Association (FA) 1863, 1866, 1869, 1871
- Running Code
  - Rugby School 1845, Rugby Union 1871

What caused the proliferation of rules after 1860? By 1842, the Gilbert football factory had moved nearer to The Close at Rugby School to be more visible to all comers, where is remains today. Further, Gilbert displayed a more regularly shaped oval ball at the Great Exhibition of 1851. Richard Lindon had lived next door to the Gilbert factory in Rugby on High Street before Gilbert moved to the current location. Lindon went into the football business and by 1862 he perfected the use of a rubber bladder and pump to replace the pig bladder and lung power. Lindon marketed a spherical ball and an oval ball. I believe that football organizers of that era knew full well the new developments in football production and that it was necessary to chose between running/tackling (using the oval ball) and kicking/dribbling (using the spherical ball). The FA rules of 1863 deleted the fair catch and "hacking" a tripping move. In 1866, the FA dropped the Eton Rouge, eliminating all in-goal activity and completely eliminating handling except by the goalkeeper. A number of clubs still employed the Sheffield Rules which allowed batting the ball, allowed pushing the ball with the body, had no offside rule and had a unique scoring system, the Sheffield Rouge, that still exists in Australian Rules and Gaelic football as will be noted later. Three side-by-side boxes (goals) were formed by four uprights with a crossbar across all four uprights. A shot into the middle box was scored as a goal while a shot into an outer box was scored as a rouge. If the game was tied on goals, the team with the most rouges was the winner. The Rugby Union formed in 1871 as the running code. By 1877 the last Sheffield Rules teams had joined the FA leaving Association football as the only kicking code.

## 3. WILLIAM WEBB ELLIS

The Rugby Union World Championship trophy is called the Webb Ellis Cup. Famously, a statue on The Close in Rugby states "With a fine disregard for the rules of football as played in his time, Webb Ellis first took the ball in his arms and ran with it thus originating the distinctive feature of the rugby game." What we actually know about Webb Ellis is that he was born at either Salford or Manchester in 1806, attended Rugby School from 1816-1825, was known as taking unfair advantage at cricket, went on to play cricket at Cambridge, became an Anglican clergyman and died in 1872.

William Bloxam wrote letters to *The Meteor* in 1876 and 1880 (four and eight years after Ellis' death) stating the only known case for Webb Ellis "William Webb Ellis whilst playing Bigside at football in 1823 caught the ball in his arms. According to the then rules he ought to have retired back … for it was by means of these placed kicks that most goals were kicked. Ellis disregarded this rule…rushed forward with the ball in his hands, with what result as to the game I know not, neither do I know how this infringement … was followed up, or when it became .. a standing rule." Bloxam mentioned no actual witness (he wasn't there himself) nor did any witness ever come forward. Webb Ellis may have cheated a few steps while taking a fair catch or kicking for an Eton Rouge, but that hardly created a new game.

An inquiry by the Old Rugbeians in 1895 states "Running with the ball in the latter half of 1823 by Mr. W. Webb Ellis …was regarded of dubious legality for some time, and only gradually became accepted as part of the game." An exhibit at the Gilbert Museum states that by 1830-1840 running in was tolerated. The ball could be kicked out for a fair catch and try at goal. It wasn't until 1841-1842 that running the ball was completely legalized, resulting in a try at goal. If Ellis did run with the ball, it took several years for any changes to be made. It should be noted that 1895 was the year when Rugby League separated from Rugby Union. Rugby Union needed to maintain visibility with the public and Webb Ellis fit the bill.

## 4. CANADIAN AND AMERICAN CODES BEGIN

Rugby reached Canada in 1861. In 1869, the first American football game was played between Rutgers and Princeton. The organizers of that game chose a combination of the two dominant kicking (non carrying) codes of the day, FA Rules and Sheffield Rules, the latter allowing batting the ball and pushing the ball forward with the body. For this rather rough kicking game, scoring was via goals kicked under a crossbar. As with the Eton-Rugby rivalry in England, rivalry between Princeton and Harvard proved important to the evolving American code. While most games from 1869 to 1875 were played under FA-Sheffield rules, Harvard preferred the "Boston Game" in which the ball could be carried if the player was pursued. Harvard played McGill of Canada in 1874 and 1875, half the game played under Boston Rules with limited carrying and half the game played under Rugby Union rules with unlimited carrying. Since Harvard preferred the Rugby Union rules, Harvard convinced American schools to switch to Rugby Union rules where goals were scored by kicking over the crossbar. That mode of play was followed from 1876 to 1882.

## 5. SCORING RULES EVOLVE

As in Table 1, scoring rules were created for the American and Canadian codes in 1883 and for Rugby Union in 1886. The evolution of scoring appears interrelated. At first, a conversion counted more than a try or touchdown, but then the conversion became lower scoring soon after. A converted touchdown/try evolved to six points and then to seven points in all three codes. Scoring by drop goals, field goals and penalty goals evolved to three points in all three codes.

Prior to 1883 a player in the three codes could catch a ball inside the 25 yard line, run back, touch the ball down in the player's own in-goal area for a safety and then be allowed to kick away from the 25 yard line. From 1876-1882 in the American code, if a game was tied on goals, the team with fewer safeties won. In the American and Canadian codes starting in 1883/1884, if a team was responsible for a ball being in its in-goal area and the ball was touched down, that constituted a safety, the other team receive two points and the ball had to be kicked to that other team also. The safety rule was intended to reduce delaying tactics.

| | Try/Touchdown and Conversion | | | | | | Kick Scoring | | | | Safety | | Rouge |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | R. Union | | Amer. | | Canad. | | R. Union | | Amer. | Canad. | Amer, | Canad. | Canad. |
| Year | Try | Conv. | TD | Conv. | TD | Conv. | Drop Goal | Pen. Goal | Field Goal | Field Goal | | | |
| 1883 | | | 2 | 4 | 4 | 4 | | | 5 | 6 | 1 | 2 | 1 |
| 1884 | | | 4 | 2 | | | | | | | 2 | | |
| 1886 | 1 | 4 | | | | | 3 | | | | | | |
| 1887 | | | | | | | | | | 5 | | | |
| 1888 | 1 | 2 | | | | | | 2 | | | | | |
| 1890 | | | | | 4 | 2 | | | | | | | |
| 1891 | 2 | 3 | | | | | 4 | 3 | | | | | |
| 1893 | 3 | 2 | | | | | | | | | | | |
| 1897 | | | 5 | 1 | | | | | | | | | |
| 1903 | | | | | | | | | | 2 | | | |
| 1904 | | | | | 5 | 1 | | | 4 | | | | |
| 1905 | | | | | | | | | | 3 | | | |
| 1906 | | | | | | | | | | 4 | | | |
| 1908 | | | | | | | | | | 3 | | | |
| 1909 | | | | | | | | | 3 | | | | |
| 1912 | | | 6 | 1 | | | | | | | | | |
| 1948 | | | | | | | 3 | | | | | | |
| 1956 | | | | | 6 | 1 | | | | | | | |
| 1958 | | | 6 | 1 or 2 | | | | | | | | | |
| 1971 | 4 | 2 | | | | | | | | | | | |
| 1975 | | | | | | 1 or 2 | | | | | | | |
| 1992 | 5 | 2 | | | | | | | | | | | |

Table 1. Evolution of Scoring in Rugby Union, American Football and Canadian Football

The other scoring rule in Table 1 is of significant historical interest. The Eton Rouge has existed under the name "Rouge" in Canadian football since 1883. If a defending team obtains a ball kicked or run into its in-goal area by an attacking team and if the defending team touches the ball down, the attacking team gets a one-point rouge but the defending team is allowed to kick the ball away, its reward for that maneuver.

## 6. RUGBY LEAGUE, DOWNS, AND TACKLES EVOLVE

Professional football began in the American code in 1895, the same year that the professional Rugby League separated from the Rugby Union in northern England. Rugby League has maintained an independent scoring system ever since, as befits the intention of striking out on its own. Under current rules, a try counts four points, a conversion two points, a penalty goal two points and a drop goal one point.

The American, Canadian and Rugby League codes have evolved systems of play continuity that replaced rucks and malls in Rugby Union. In the American and Canadian codes, downs are the number of plays required to make a yardage target. If that target is met, additional downs are earned. In Rugby League the target is to score in the given number of tackles. Here is the evolution of downs and tackles

- American Football
  1882 3 downs to earn for 5 yards
  1906 4 downs to earn 10 yards
- Canadian football
  1887 heeling back is allowed
  1903 3 downs to earn 10 yards
- Rugby League
  1906 play the ball/heeling is allowed
  1966 4 tackles to score
  1971 6 tackles to score

## 7. OFFSIDE RULES

In Rugby Union, Rugby League, American football and Canadian football, offside rules depend generally on the position of the ball under certain playing conditions. This contrasts with current Association football offside rules that relate to the position of certain numbers of defenders at the moment the ball is played. Earlier, offside rules for Eton (1847) and Cambridge (1856) required four defenders. The 1866 FA rules called for three defenders while the 1925 FA rules lowered that to two defenders. In 1990, FIFA made the two-defender rule less strict in that one defender can be even with the player attempting to receive the call.

## 8. THE FORWARD PASS IS LEGALIZED

American football had become such a rough sport by 1905 that 18 died playing the sport that year. Players would link arms in front of a ball carrier requiring an opposing player to hurtle into the “flying wedge”. With three downs to progress five yards, a ball carrier would be thrown forward over attackers. Then President Theodore Roosevelt threatened to outlaw football unless the game was made safer. A group of American colleges formed what is now called the National Collegiate Athletic Association and that group formed a football rules committee. The flying wedge was outlawed, teams were separated by a “line of scrimmage”, four downs were used to earn ten yards and the forward pass was legalized. The first legally-completed forward pass was thrown in September 1906 from quarterback Brad Robinson of Saint Louis University to receiver Jack Schneider. Thus began what has become the most identifiable part of American football. Canadian football legalized the forward pass in 1929.

## 9. AUSTRALIAN RULES AND GAELIC FOOTBALL

A statue outside the Melbourne Cricket Grounds commemorates what is said to the first Australian Rules football game played in 1858 between Scotch College and the Melbourne Grammar school. Prominent in the statue is Tom Wills, a referee for that match. Wills was one of the principals in creating that code. To understand the mind set of Wills, a brief biography is useful. Wills was born in 1835 in New South Wales, grew up with Aboriginals and spoke an Aboriginal dialect. In 1850 he was sent to England to be educated at Rugby School. He later played cricket at Cambridge. Upon his return to Australia, he organized a tour by an Aboriginal cricket team. He and others created what was to be “A game of our own”.

Wills would have known of various football codes in England and would have known of an Aboriginal sport called Marn Grook in which a stuffed possum skin was kicked into the air. Players would leap into the air and catch the skin. The written 1859 Geelong and Melbourne Rules allowed batting the ball, kicking the ball and pushing the ball forward with no offside rule, all features of Sheffield Rules. The 1866 rules introduced two additional vertical “behind” goal posts, one on either side of the vertical main goals posts. That created use of the Sheffield Rouge, described earlier. Under current scoring values, a kick through the middle posts scores a six-point goal while a kick through either set of outer posts scores a one-point behind, basically a Sheffield Rouge. The ball could be carried forward as in Rugby rules. I believe that rules makers purposely chose Sheffield Rules to create a non-London/FA game with Rugby rules thrown in to make the game unique. An oval ball was chosen. The fair catch/mark was also present in the Australian Rules code, as it was in all codes as of 1858, not to be deleted from FA rules until 1863, and in Marn Grook. It is not possible to know whether the fair catch was included to be consistent with other known codes or whether it was intended to follow Marn Grook. Whatever the actual intention, the quintessential feature of Australian Rules football, the high-flying mark, is coincidental to an Aboriginal game.

By 1888, the Gaelic Athletic Association published rules for Gaelic football that were nearly identical to those of Australian Rules football. I believe that both set of rules were purposely based on Sheffield Rules so as to be independent of London influence. There were earlier Gaelic football games but no written rules exist. Lacking such rules, it cannot be ascertained whether Irish emigrating to Australia brought their game and influenced Australian Rules football or that Irish returning to Ireland from Australia changed Gaelic football. A spherical ball is used with Rugby-styled, T-shaped goal posts. A kick under the cross bar against a goalkeeper scores a three-point goal while a kick over the goal ports scores one point, the counterpart to the Sheffield Rouge. International Rules have been created as a hybrid of Australian and Gaelic codes, allowing for matches between those two countries.

**10. Summary**

The seven football codes of today have interrelated evolutionary histories. Through the early 1800s, there was one rough game with few rules but with one ball: a pig bladder later covered with a leather casing. Two paths emerged, one with two branches. The public path involved flexible rules and the Sheffield Rouge, played by pubs and clubs. That path led to **Australian Rules football** in 1857 and to **Gaelic football** in 1888. The schools path, invoking stricter rules and sportsmanship, has two branches. The Eton-influenced branch preferred a kicking game with goals scored by kicking under a crossbar, leading to Football Association rules (**Soccer**) in 1863 and the first games played in American football from 1869 to 1875. The other schools branch was influenced by Rugby School involving running, with goals kicked over a cross bar. That branch led to **Canadian football** rugby in1861, the **Rugby Union** in 1871, **American Football** involving rugby in 1876 and **Rugby League** in 1895. Scoring rules evolved and rules changes evolved to create today's codes.

Three important people are connected with the evolution chronicled above. Rules emerged largely after 1862 when Richard Lindon invented a rubber bladder, resulting in a spherical ball and an oblong ball, requiring a conscious decision to be made between a kicking code and a running code. William Webb Ellis was more of a convenient marketing tool thsn a (supposedly) audacious pioneer, when in fact he may only have cheated a few steps taking a mark in 1823; an act only mentioned by Matthew Bloxam in 1876, more than 50 years after the fact and four years after his death. Australian Tom Wills attended Rugby School and shared a skill at cricket with Webb Ellis. Wills helped create a unique Australian Rules game which probably spread to Ireland by Irish who had lived in Australia.

In our day, the codes have come to be described by words from the former All Blacks of New Zealand haka, "*Ka ora, ka mata*", that is, they are a matter of "life or death" to fans.

**References**


Alcock, C.W. (1906) *Football: The association game*. London:Bell.

Davis, P. The Beginning of Football at Harvard, *Harvard Alumni Bulletin*, September 29, 1915, 213-220.

en.wikipedia.org/wiki/Eton_field_game

en.wikipedia.org/wiki/History_of_American_football

en.wikipedia.org/wiki/History_of_Australian_rules_football_in_Victoria_(1859-1900)

en.wikipedia.org/wiki/Rugby_league_gameplay

en.wikipedia.org/wiki/Sheffield_Rules

en.wikipedia.org/wiki/The_Cambridge_Rules

en.wikipedia.org/wiki/Tom_Wills

en.wikipedia.org/wiki/Two-point_conversion

Fabian, A.H. and Green, G., eds., (1960) *Association football*, Vols. 1-4, London:Caxton.

Green, G. (1953) *The history of the football Association*. London:Naldrett.

Harvey, A. (2005) *Football: the first hundred years: the untold story*. Routledge: Uk.

Jackson, N.L. (1900) *Association football*. London:Newness.

Lecke, R. (1973) *The Story of (American) Football*. Random House:New York

Orejan, J. (2006) The history of Gaelic football and the Gaelic Athletic Association, *The Sport Management and Related Topics Journal*, 2(2), 45-48..

Rugby: www.rugbyfootballhistory.com/originsofrugby.htm

Sproule, B. (1991) Canadian Football: Past to Present. *The Coffin Corner*, 13(1), 1-5.

www.richardlindon.com

www.rugbyfootballhistory.com/bloxam.html

www.rugbyfootballhistory.com/laws.htm

www.scarletknights.com/football/history/first-game.asp

www.the-game.org/history-originto1889.htm

# Determination of Influential Variations for Rating Australasian Greyhounds

Anthony Bedford[a,b], Joshua Kadlec-Cavanagh [b] and Shane Clarke [b]

[a] *School of Mathematical and Geospatial Science*

*RMIT University, Melbourne, Australia*

[b] *XTrade, Australia and Hong Kong*

*Corresponding author: Anthony.Bedford@rmit.edu.au*

**Abstract**

In this research, we investigate the difference in track variations for greyhounds in Australia and New Zealand. The purpose is to ascertain what expectations are required to win at certain tracks, and a way of discriminating ratings for dogs when they change tracks. Emergent patterns in variations are also clear when class of race is considered. A number of interesting findings do emerge, and one needs to be careful in catering for a variety of confounding variables. There exists a day effect, a function of the class of races on those days; turns have a strong influence; distance is more obvious; geography; turf type and quantity of meetings rounds out the variables of influence. Analysis is split by track, distance and class, accounting for day effects. The key to this work is the setting of a universal basis to generate ratings for inter and intra course relevancy. We exhibit our interface for trading, and how having critical information at hand makes the traders life simple.

*Keywords*: Greyhounds, prediction, gambling, racing

# ADJUSTING TRUE ODDS TO ALLOW FOR VIGORISH

Stephen R Clarke
*Swinburne University of Technology*
sandkclarke@hotmail.com

## Abstract

A mathematical model predicting sporting outcomes produces probabilities that sum to one, whereas the probabilities implied by bookmaker's odds sum to more than one. Vigorish or over-round is the excess probability that supplies the bookies margin, and could typically range from 2% to over 20%. If the probabilities from a mathematical model are to be used for supplying real time odds they need to be adjusted for the bookmaker's vigorish. In the reverse situation, when testing models from past data, bookmaker's odds need to be adjusted so that the implied probabilities sum to one. Schembri et al (2011) discusses two methods – normalisation and equal distribution. However neither of these suitably allow for the fact that the margin on outsiders is usually greater than favourites. A true price of $1.05 can be reduced to $1.03 for less than 2% margin, whereas a $100 true price could be set at $50 for a $50% margin. This paper discusses an alternative approach using a power function to transform probabilities. This was successfully used when supplying real time odds to a leading bookmarker (Clarke, 2007).

**Keywords**: Betting, over-round

### 1. INTRODUCTION

In some sporting studies we need to estimate the chances of past or future events. Often the only estimates available are historical bookmaker's odds or prices. When these are converted to probabilities they sum to more than one. This excess probability is known as the over-round O. This supplies the bookies margin since it results in payouts that are less than justified by the true probabilities. An alternative measure used is the vigorish V, or proportion of the amount bet that the bookmaker retains from a balanced book. Thus V = O/ (1 + O), which in casino applications is called the house percentage. In order to estimate the true probabilities the over-round or vigorish needs to be removed so the probabilities sum to one. In this paper we also use R, the expected proportion of amount bet that is returned to the punter, where R = 1 – V.

Two methods have generally been used in the literature. Schembri et al (2011) and Viney (et al) discuss two methods – normalisation and equal distribution. Equal distribution merely subtracts an equal amount from each probability, whereas normalisation reduces each probability by the same proportion. Thus for two outcomes and implied probabilities of P and Q (P+Q > 1, so O =P+Q -1) we have:

Equal distribution results in p = P - O/2, q = Q - O/2

Normalisation results in p = P / (P+Q), q = Q / (P+Q).

So for example in a tennis match where the prices are $4.04 and $1.20, giving implied probabilities of 0.25 and 0.83 for an over-round of 8% (V = 7.5%, R = 92.5%), equal distribution gives true probabilities of 0.21 and 0.79 (fair prices of $4.83 and $1.26), whereas normalisation results in 0.23 and 0.77 (fair prices of $4.37 and $1.30).

There are some problems with such approaches. The normalisation method results in probabilities that are reduced by a common percentage of V (7.5% in the above example). ie the same percentage is taken from all bets. However it is common practice for bookmakers to take a greater margin out of longer priced outcomes. Clearly the equal distribution method achieves this, with the outsider's probability reduced by 16% and the favourite by only 5%. However this often goes too far, and can produce negative probabilities, particularly when there are several outcomes. For example a horse race with 5 runners at $40, $30, $20, $5 and a hot favourite at $1.12 results in an over-round of 20%, which gives non-positive probabilities for the 3 outsiders.

This paper looks at an alternative. While equal distribution is an additive model and normalisation a multiplicative model, here we discuss a power model.

### 2. ADJUSTING TRUE PROBABILITIES TO ALLOW FOR VIGORISH.

This problem first arose when providing real time odds for a betting company. A regression model was used to produce estimated probabilities for the number of runs scored in an over of cricket. These had to be adjusted to allow for the vigorish required by the bookmaker, and the solution to take an equal percentage off all prices was considered inadequate. If we wish to return only a proportion R of the amount bet to punters, we need to reduce the payout for an event with probability p and a fair price of 1/p to R/p. Since this return must be greater than 1, it could only be viable when p <R. So for example, for a return to punter of 80% (house

percentage 20%) the shortest price we could adjust would be 1/.8 = $1.25 or 4 to 1 on. Roulette can use this system to decrease all fair prices by the same percentage only because it has a high R (97% or 94%) and the shortest price bets are even money.

Unlike roulette, it is common to take a higher percentage from low probability/high payout events. Thus a 1000 to one shot can be given a payout of $500 for a return to punter of only 50%, while a hot favourite with a true price of $1.02 can barely be reduced at all.

One way to implement this is to give a punter the same return for a double as betting on the two individual events.

Let the payout for an event with probability x be P(x).

Then if winnings are placed all up on a bet with probability y, the final payout is P(y) P(x)

Alternatively, since the double has a probability of xy, the payout on that will be P(xy).

So we want P(xy)=P(x)P(y), and the function satisfying this is the power function $P(x) = 1/x^k$ and we need k < 1 so the price is reduced, not increased.

Thus while equal distribution alters probabilities by an additive constant, normalisation by a constant multiplier, the power method raises them by a constant power. k depends on the return to gambler R. Taking logs we get k = -log(P) / log(x). Thus for a 50/50 bet with a return to punter R the payout will be 2R so k = log(2R)/log2. (For a bet with n equally likely outcomes, k = log (nR)/log n )

So for example if the return for a 50% bet is R = 90.0%, payout is 2R= 1.8, and k = log(1.8)/log2 = 0.848. While this is only exact for 2 equally likely outcomes, it can be used as an approximation. Table 1gives the adjusted probabilities and prices obtained using this method for a range of true probability events. The expected return to punters is for any bet is $xP = x.(1/x^k) = x^{1-k}$, equal to 90% for a 50% bet, higher for favourites and less for outsiders.

| True Probability x | Fair Price 1/x | Adjusted Probability $x^k$ | Adjusted Price $1/x^k$ | Expected Return $x^{1-k}$ |
|---|---|---|---|---|
| 0.01 | $100.00 | 0.02 | $49.66 | 50% |
| 0.05 | $20.00 | 0.08 | $12.68 | 63% |
| 0.1 | $10.00 | 0.14 | $7.05 | 70% |
| 0.2 | $5.00 | 0.26 | $3.91 | 78% |
| 0.25 | $4.00 | 0.31 | $3.24 | 81% |
| 0.3 | $3.33 | 0.36 | $2.78 | 83% |
| 0.4 | $2.50 | 0.46 | $2.17 | 87% |
| 0.5 | $2.00 | 0.56 | $1.80 | 90% |
| 0.6 | $1.67 | 0.65 | $1.54 | 93% |
| 0.7 | $1.43 | 0.74 | $1.35 | 95% |
| 0.8 | $1.25 | 0.83 | $1.21 | 97% |
| 0.9 | $1.11 | 0.91 | $1.09 | 98% |
| 0.95 | $1.05 | 0.96 | $1.04 | 99% |
| 0.99 | $1.01 | 0.99 | $1.009 | 99.8% |

Table 1: Adjusted probabilities and prices using the power method with k = 0.848

With 11 outcomes k = log(11R)/log(11) =0.932 for R = 85% and 0.907 for R = 80%. Table 2 shows the resultant prices for a range of bets for a nominal R = 85% and 80%, along with the actual expected percentage return to the punter. Note the returns are close to the expected values for values around the average $11.00 payout. In our application it was felt the adjusted prices obtained were realistic, and the returns were close enough to that expected for the formula to be used in real time rather than many tables for varying R.

| Fair | Nominal R = 85% (k= 0.932) | | Nominal R = 80% (k=.907) | |
|---|---|---|---|---|
| Payout | Payout | % return | Payout | % return |
| $P$ | $P^k$ | $P^{k-1}$ | $P^k$ | $P^{k-1}$ |
| $40.00 | $31.15 | 78% | $28.38 | 71% |
| $33.33 | $26.28 | 79% | $24.05 | 72% |
| $28.57 | $22.76 | 80% | $20.91 | 73% |
| $25.00 | $20.10 | 80% | $18.53 | 74% |
| $22.22 | $18.01 | 81% | $16.65 | 75% |
| $20.00 | $16.32 | 82% | $15.13 | 76% |
| $18.18 | $14.94 | 82% | $13.88 | 76% |
| $16.67 | $13.77 | 83% | $12.83 | 77% |
| $15.38 | $12.78 | 83% | $11.93 | 78% |
| $14.29 | $11.93 | 83% | $11.15 | 78% |
| $13.33 | $11.19 | 84% | $10.48 | 79% |
| $12.50 | $10.53 | 84% | $9.88 | 79% |
| $11.76 | $9.95 | 85% | $9.35 | 80% |
| $11.11 | $9.44 | 85% | $8.88 | 80% |
| $11.00 | $9.35 | 85% | $8.80 | 80% |
| $10.53 | $8.97 | 85% | $8.46 | 80% |
| $10.00 | $8.56 | 86% | $8.07 | 81% |
| $6.67 | $5.86 | 88% | $5.59 | 84% |
| $5.00 | $4.48 | 90% | $4.30 | 86% |
| $4.00 | $3.64 | 91% | $3.52 | 88% |
| $3.33 | $3.07 | 92% | $2.98 | 89% |
| $2.86 | $2.66 | 93% | $2.59 | 91% |
| $2.50 | $2.35 | 94% | $2.30 | 92% |
| $2.22 | $2.11 | 95% | $2.06 | 93% |
| $2.00 | $1.91 | 96% | $1.88 | 94% |

Table 2: Adjusted prices and expected returns for an 11 outcome event

**3. ACTUAL HOUSE PERCENTAGES OBTAINED IN PRACTICE.**

Because the house percentage of different bets changes, the overall percentage taken depends on the distribution of the probabilities of the particular outcomes, and the amounts bet. Thus for example, we expect an event where there are equally probable outcomes to have a different return than one in which there are one or two hot favourites and the rest are highly unlikely.

Expected return to punter = (Σ bet size * true prob winning * payout )/(total bet)

= (Σ proportion of pool bet * true prob winning * payout )

If punters bet in the same proportion as the probabilities, we have,

Expected return to punter = Σ x*x* payout, where x is the probability of an outcome

(Note for the constant percentage case, payout = R/x, so expected return to punter = R as required)

Using the power formula, payout = $1/x^k$, so expected return = $\Sigma x^{2-k}$

It is easily shown using Lagrange Multipiers that the maximum value of this is R when all x's are equal.

So assuming n outcomes all with equal probability of 1/n we finally get

Maximum expected return = $\Sigma x^{2-k} = \Sigma(1/n)^{2-k} = n.(1/n)^{2-k} = (1/n)^{1-k} = R$, since

$k = \log(nR)/\log(n) = 1 + \log(R)/\log(n)$ , so $\log(R) = (k-1) \log(n) = \log(1/n)^{1-k}$

Thus the value used for R is in general a maximum and actual returns will be less than this.

In the two outcome example this means the target return is only achieved for two equal opponents. Table 3 shows the true and reduced prices and the expected return to the bookmaker for a balanced book on a two outcome event for a sample of markets. The return to the bookmaker only deviates markedly from the expected percentage if the outsider is less than a 30% chance.

| Probabilities of 2 outcomes | | True Prices | | Reduced prices | | Return to Bookmaker |
|---|---|---|---|---|---|---|
| 0.5 | 0.5 | $2.00 | $2.00 | $1.80 | $1.80 | 10.0% |
| 0.4 | 0.6 | $2.50 | $1.67 | $2.17 | $1.54 | 9.8% |
| 0.3 | 0.7 | $3.33 | $1.43 | $2.78 | $1.35 | 9.0% |
| 0.2 | 0.8 | $5.00 | $1.25 | $3.91 | $1.21 | 7.7% |
| 0.1 | 0.9 | $10.00 | $1.11 | $7.05 | $1.09 | 5.3% |
| 0.02 | 0.98 | $50.00 | $1.02 | $27.59 | $1.02 | 1.9% |
| 0.01 | 0.99 | $100.00 | $1.01 | $49.66 | $1.01 | 1.2% |

Table 3: Return to bookmaker for a balanced book on a 2 outcome event using Power method to convert true probabilities with a nominal vigorish of 10%.

**4. ADJUSTING BOOKMAKERS ODDS TO ALLOW FOR VIGORISH.**

The above shows we can use the formula $P = 1/x^k$ to adjust prices of events with true probability x to allow for a return to punter of R, where k = log (nR)/ log(n). This gives adjusted probabilities that give a return to the punter of exactly R when all events are equally likely, but less than this in other cases.

When used in reverse to remove over-round from bookmakers prices P we have adjusted price = 1/x = $P^{(1/k)}$ or Adjusted price = $P^{k'}$ where k' = log(n)/log (nR) where n is the number of outcomes and R is the return to punter = (1- O). However since this only gives the required R for equally likely outcomes, we need to use iteration to produce probabilities that sum to 1. This is easily performed in a spreadsheet.

Consider the prices for a tennis match where published prices are $1.22 and $4.33. This gives implied probabilities of 0.820 and 0.231 for an over-round of 0.051 and return to punter R of 0.952 and so k = 0.929. The equal probability method distributes the 0.051 equally for probabilities of 0.794 and 0.206 (prices of $1.26 and $4.88). The normalisation method takes each probability as a proportion of the total for probabilities of 0.780 and 0.220 (prices of $1.28 and $4.55) Note this is equivalent to increasing each price by the (required) same proportion. The power method initially gives probabilities of .807 and 0.206, but these still sum to more than one. Using iteration to adjust k to correct this, we obtain k = 0.904 and probabilities of 0.802 and 0.198 (or prices of $1.25 and $5.05). Clearly this method increases the prices of the outsiders to a greater extent than the other two methods.

In events with a larger number of competitors such as horse racing, the outsiders are at longer odds and the over-rounds are much greater than in two person events such as tennis. Table 4 shows the three methods applied to a race with 6 runners.

| Prices and their Implied probabilities | | Calculated True Probabilities: Equal Distribution | Calculated True Probabilities: Normalisation | Calculated True Probabilities: Power | Calculated Fair Prices: Equal Distribution | Calculated Fair Prices: Normalisation | Calculated Fair Prices: Power |
|---|---|---|---|---|---|---|---|
| $1.15 | 0.870 | 0.828 | 0.696 | 0.825 | $1.21 | $1.44 | $1.21 |
| $5.00 | 0.200 | 0.158 | 0.160 | 0.110 | $6.31 | $6.25 | $9.12 |
| $10.00 | 0.100 | 0.058 | 0.080 | 0.042 | $17.12 | $12.50 | $23.63 |
| $20.00 | 0.050 | 0.008 | 0.040 | 0.016 | $118.97 | $24.99 | $61.23 |
| $50.00 | 0.020 | -0.022 | 0.016 | 0.005 | -$46.31 | $62.48 | $215.54 |
| $100.00 | 0.010 | -0.032 | 0.008 | 0.002 | -$31.65 | $124.96 | $558.47 |
| Total | 1.250 | 1.000 | 1.000 | 1.000 | | | |

Table 4: Comparison of 3 methods of adjusting prices to remove 20% vigorish

With an over-round of 25% or return to punter of 80% the equal probability method breaks down giving negative probabilities for the outsiders. The normalisation method merely reduces all probabilities by 20% (increases prices by 25%). Schembri et. al. (2011) concluded that the normalisation method is less effective when there is a strong favourite, as too much over-round is given to the favourite. The iterative power method with an initial k of 0.876 iterates to k = 0.728 adjusts the favourites to a lesser extent than the normalisation, but adjusts the outsiders much more. The power method thus avoids the problems the equal distribution method has with outsiders, and the over allocation the normalisation method has with favourites.

## 5. CONCLUSION

A power method can be used to adjust true probabilities to ones that sum to more than one to allow for over-round. Used in reverse requires iteration. The method should be considered, as it more truly allows for the practice of taking a greater percentage out of winning bets on outsiders than favourites.

**Acknowledgements**

This work originally arose from a project working with Mark Solonsch, and resulted from discussions I had with him on adjusting odds.

## References

Clarke, S.R. (2007) Successful applications of statistical modeling to betting markets. In IMA Sport 2007: First International conference on Mathematics in Sport. D. Percy, P Scarf & C Robinson (Eds.) The Institute of Mathematics and its Applications, Salford, UK. pp35-43.

Schembri, A., Bedford, A., O'Bree, B. & Bressanutti, N. (2011). Comparing market efficiency with traditional and non-traditional ratings systems in ATP tennis. *Journal of Quantitative Analysis in Sports* (JQAS),1.

Viney, M., Bedford, A. & Kondo, E. (2013). Incorporating over-round into in-play Markov Chain models in tennis. 15th International Conference on Gambling & Risk-Taking, Las Vegas, USA.

# IN-PLAY BETTING AS DYNAMIC PORTFOLIO RISK MANAGEMENT

Darren O'Shaughnessy [a,b,c]

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

## Abstract

The Kelly criterion (Kelly, 1956) is well known as the optimal path to bankroll growth when exploiting probabilistic knowledge of future outcomes. Log-utility maximisation is an exercise in risk management performed in portfolio management as well as sports betting. O'Shaughnessy (2012) and Noon (2014) extended this to a solution in win-draw-loss markets involving both bet and lay options.

With the explosion in in-play betting now accounting for over 75% of volume in European markets (Munro, 2016), there is a need for fast algorithms that manage the bettor's current position across several different correlated markets in each game.

In high-scoring sports such as Australian Rules Football and basketball, the "line" markets where one team is given a handicap move dynamically both before and during play, while the odds (or prices) on each side of that line stay relatively constant. A market in which the bettor previously invested – e.g., San Antonio −8.5 – may cease to exist as the popular "money line" is adjusted, so it cannot be hedged directly.

This paper derives a general mathematical expression of the quantity to maximise as a function of past bets and multiple live markets, and uses the multi-dimensional version of the Newton-Raphson method to quickly identify the correct mix of bets in the currently available markets. This involves (relatively simple) calculation of the first and second partial derivatives of the generalised utility function with respect to variables representing each market type.

*Keywords*: Kelly criterion, risk management, Australian Rules Football, AFL, basketball, Newton-Raphson method

## 1. INTRODUCTION

Investors have recognised since Kelly (1956) that to maximise the growth rate of their bankroll in an uncertain market, they must apportion an amount based on optimisation of log-utility. For more background regarding the extended development of this criterion in modern betting and investment markets, please see Noon (2014), MacLean, Thorp & Ziemba (2010), and O'Shaughnessy (2012).

Most published research in this area deals with one moment of decision-making per match, in a largely static set of markets. Once the bet has been placed, and the odds available change, further profit or risk-management is available by taking advantage of the new prices in association with the bettor's current market profile. The simplest example of this is arbitrage, where a punter can guarantee a profit by betting on the complement of his existing bet with a finely tuned amount at an increased price; however this is quite different to probabilistic utility maximisation, where a partial hedge is more likely to be the optimal strategy.

Despite the anomalous legal status of in-play betting[11] in Australia, bookmakers report increasing interest and turnover from punters watching the game and adjusting their bets live. In some European sporting markets where it is legal to offer in-play markets on the internet, 75% of all liquidity is now wagered after the sporting event has commenced (Munro, 2016).

In addition to dynamic odds, the punter is confronted with changing margins. In high-scoring sports such as basketball and Australian Rules football, the two most liquid markets in practice are the head-to-head odds (H2H) and a "points" line (Line) with roughly even odds, which represents the median expected winning margin. These are the most easily understood by watchers of the game, and usually have the lowest overround in the bookmaker's portfolio. The bookmaker generally offers other fixed handicaps simultaneously, such as for one team to win by 40 or more points. As the main Line market is updated during the game, if a bettor keeps placing bets he can develop a complex portfolio of profits and losses on ranges of scores. For example, he may have active bets on Team A to win, Team B to not lose by more than 15.5 points, and Team A to win

[11] Bookmakers registered in Australia may legally accept bets after the event has started, but only in person or via a voice call, not electronically over the internet. This legislation is unique globally, intended to curb problem gambling. It also has the effect of suppressing Australian bettors' exploitation of changing odds and knowledge.

by more than 21.5 points, while the new market offers a \$1.90 price of either team's side of a 10.5 point Line and \$3.75 for Team B to win. Maximising log-utility was never trivial, and it now requires a sophisticated formula and algorithm.

This paper derives that formula and shows that the relatively simple Newton-Raphson method is guaranteed to find a solution if it exists. Eklund (2011) previously applied the two-dimensional Newton-Raphson method to bookmaker's markets; this paper extends that to multiple dimensions with each dimension representing the variable bet size in an available market.

## 2. METHODS

### KELLY CRITERION AND LOG-UTILITY

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

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

where $M$ is the team's market price and $p$ is the gambler's presumed probability of the team winning. $\beta$ is expressed as a percentage of the bettor's bankroll, and a bet should be placed if $Mp>1$. The formula is derived by maximising the expected value of $W = log$(bank) with respect to the bet size $x$ in equation (2):

$$W = (1-p)\log(B-x) + p\log(B+(M-1)x) \qquad (2)$$

where $B$ is the initial bankroll. The solution to maximise $W$ here is $x = \beta B$.

In order to develop the general expression for $W$ in a dynamic environment with multiple markets available, we partition the result landscape into multiple margin ranges. Any bet that has already been taken, or is available, or leads to a changed payout, is included as a boundary of one of these partitions. For example, in the example in the introduction, we calculate probabilities $p_i$ for each of these scenarios:

1. Team A wins by more than 21.5 points
2. Team A wins by between 15.5 – 21.5 points
3. Team A wins by between 10.5 – 15.5 points
4. Team A wins by between 0.5 – 10.5 points
5. A draw (bookmaker pays all H2H bets at half value)
6. Team B wins

The nature of this probability calculation is beyond the scope of this paper, but it would generally involve Monte Carlo simulation or an estimation of path evolution from the current state of the game to final scores. In practice, to counter the risks of the punter overbetting using the Kelly formula (MacLean, Thorp and Ziemba, 2010), he should include the public's knowledge as expressed in the market. A Bayesian approach is preferred to relying 100% on his own flawed model.

Bets on markets which are not predicated on the final margin are not considered here, although there can be value (and interdependent risk management) available in markets such as half-time margins. Nor are "index" bets that scale with the size of the difference from the final result considered, although they could be brought into this framework.

In each partition $\underline{i}$, we calculate the static net payout $s_i$ from all existing bets, which are known for that margin range. For each available market, we test whether the smallest possible bet would lead to a marginal increase in utility. Those bets $x_j$ at odds $M_i$ which have a prospective positive effect are included in the formula and collated as vector variable $\boldsymbol{x}$.

This leads to the log-utility

$$W = \sum_{i,j} p_i \log(B + s_i + (\delta_{ij} M_j - 1)x_j) \qquad (3)$$

where $\delta_{ij}$ is 1 if bet $x_j$ is successful in partition $i$, 0 if bet $x_j$ is a loser in partition $i$, and may have other values such as 0.5 if the partition would result in a part-payout. If the bet is in an exchange market such as Betfair with a tax $t$, then for a winning partition $\delta_{ij} = 1 - t\,(M_j - 1) / M_j$

We now seek to maximise $W(\boldsymbol{x})$.

### NEWTON-RAPHSON METHOD

Note that each term of (3) is a concave function and therefore the sum $W(\boldsymbol{x})$ is concave (i.e., $-W(\boldsymbol{x})$ is convex with each second-derivative strictly positive $\forall x_j$). This makes maximisation especially amenable to the root-finding method of Newton and Raphson. Noon (2014) questions this convexity when the punter uses a betting exchange with multiple options, such as win-draw-loss as discussed in O'Shaughnessy (2012) which introduces a Heaviside function and therefore loses its differentiability. Noon uses a computational approximation to Heaviside which spoils the convexity, however I note that the only effect of the discontinuity in the derivative is to introduce a single instantaneous decrease in gradient of the function where it changes from loss to profit in that market, and the function $W$ remains piecewise continuous and wholly concave.

Walking through the Newton-Raphson method, we need to iterate the following algorithm until the solution is virtually stationary. The starting point $\boldsymbol{x}_c = \boldsymbol{x}_0 = \boldsymbol{0}$ seems to work well enough.

1. Construct the Jacobian $\boldsymbol{J}$ at $\boldsymbol{x}_c$ using the second-derivatives of $W$, for example in two dimensions

$$J = \begin{bmatrix} \frac{\partial^2 W}{\partial x_1^2} & \frac{\partial^2 W}{\partial x_1 \partial x_2} \\ \frac{\partial^2 W}{\partial x_1 \partial x_2} & \frac{\partial^2 W}{\partial x_2^2} \end{bmatrix} \qquad (4)$$

2. Calculate the gradient vector

$$\nabla W = \begin{bmatrix} \frac{\partial W}{\partial x_1} \\ \frac{\partial W}{\partial x_2} \\ \vdots \\ \frac{\partial W}{\partial x_n} \end{bmatrix} \qquad (5)$$

3. $\boldsymbol{x}_c' := \boldsymbol{x}_c - \boldsymbol{J}^{-1} \nabla W$
4. Eliminate any $x_i < 0$ to reduce the dimensions of the search space as we cannot bet a negative amount
5. If $| \boldsymbol{x}_c' - \boldsymbol{x}_c | < \xi$ ($\xi$ small), $\boldsymbol{x}_c'$ is accepted as the solution; otherwise $\boldsymbol{x}_c := \boldsymbol{x}_c'$

Partial derivatives of $W$ are presented here:

$$\frac{\partial W}{\partial x_k} = \sum_{i,j} \frac{p_i(\delta_{ik} M_k - 1)}{B + s_i + (\delta_{ij} M_j - 1) x_j} \qquad (6)$$

$$\frac{\partial^2 W}{\partial x_k \partial x_{k'}} = -\sum_{i,j} \frac{p_i(\delta_{ik} M_k - 1)(\delta_{ik'} M_{k'} - 1)}{[B + s_i + (\delta_{ij} M_j - 1) x_j]^2} \qquad (7)$$

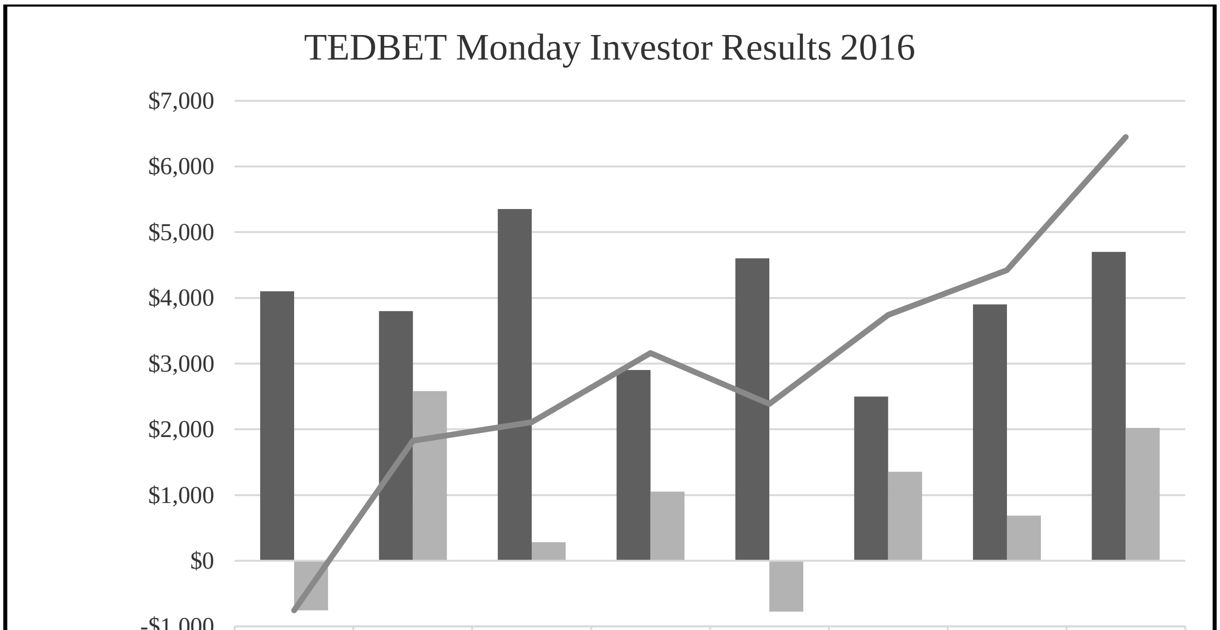
## 3. RESULTS

IMPLEMENTATION

The algorithm was implemented in PHP on a Unix webserver and runs in a live environment for all AFL matches for TedSport (see Acknowledgements). Solutions, if they exist for the two best available markets, are generally found within six iterations of Newton-Raphson and presented to the user within 100 milliseconds. The user is able to add as many H2H or Line bets as she likes and is shown her current market position.

Full profit results are not available at this time as individuals are often unable to place a real bet of the requested size with bookmakers, due to physical and time limitations imposed by the legislation. However the pre-match suggested bets using this risk-management algorithm and a forecasting model built with TedSport KPIs are currently (as at May 15th 2016) showing a 20.2% return[12] on investment (ROI), displayed in Figure 1 as designed by TedSport's Glenn McLeod:

[12] This is calculated from a constant bankroll of \$10,000, staking a total of \$31,850 across eight weeks for a profit of \$6,446. Past returns are no guarantee of future profits, and even the best systems generally regress to under 10% ROI in the long term




| | Round 1 | Round 2 | Round 3 | Round 4 | Round 5 | Round 6 | Round 7 | Round 8 |
|---|---|---|---|---|---|---|---|---|
| Staked | $4,100 | $3,800 | $5,350 | $2,900 | $4,600 | $2,500 | $3,900 | $4,700 |
| Net Result | -$754 | $2,580 | $284 | $1,051 | -$777 | $1,354 | $686 | $2,022 |
| Cumulative | -$754 | $1,826 | $2,110 | $3,161 | $2,384 | $3,738 | $4,424 | $6,446 |

Figure 1: Results of Risk Management Algorithm in AFL 2016, based on a fixed bankroll of $10,000

## 4. DISCUSSION

Many punters with excellent knowledge of their sport are unable to convert that expertise into profit due to mismanaged staking strategies. Opportunities for a prediction model in-play against bookmakers are developing, especially in a relatively isolated sport such as Australian Rules football. The ability to continuously add bets to the bettor's portfolio that increase their expected bankroll (or its utility) should be of value.

Noon (2014) develops the other side of the Kelly equation and uses it to advise bookmakers how they should set their markets, and how punters should present stakes for matching on exchanges. Potentially this paper's root-finding technique is also amenable to those optimisation problems.

It should be noted that the algorithm's staking strategy is unlike human strategies, where the punters are risk-averse – they tend to hate losing money more than they like winning it, as noted by Kahneman (2011) among others. The algorithm can manage a portfolio in-game to a guaranteed small loss for instance, in order to avoid debilitating losses. This could be a barrier to uptake, or a utility function that reflects human risk-reward behaviour could be developed.

## 5. CONCLUSIONS

This paper has presented a general formula for dynamic risk management in a live sports betting market, and demonstrated that a unique solution is calculable within a short time as the odds, scores and live information change. Results and profitability – naturally depending on a high-quality forecasting engine – are encouraging in the first few months of the implementation.

**Acknowledgements**

Part of this research was conducted under a commercial arrangement with QIQ Corporation Pty Ltd trading as TedSport, and the algorithm is implemented on the website tedbet.com.au.

**References**

Eklund, J. (2011). Discrete gambles: Theoretical study of optimal bet allocations for the expo-power utility gambler. *MSc Thesis, Stockholm School of Economics*

Kahneman, D. (2011). *Thinking Fast and Slow*. Farrar, Straus and Giroux.

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

MacLean, L.C., Thorp, E.O., and Ziemba, W.T. (2010). Good and bad properties of the Kelly criterion. *The Kelly Capital Growth Investment Criterion: Theory and Practice.* Scientific Press, Singapore

Munro, K. (2016). Bookmakers and the gambling industry push to legalise live betting. *The Age*, 10th January 2016. Available at *http://www.theage.com.au/national/bookmakers-and-the-gambling-industry-push-to-legalise-live-betting-20160107-gm18s8.html*

Noon, E. (2014). Extending Kelly Staking Strategies to Peer-to-Peer Betting Exchanges. *PhD Thesis, Imperial College London, Department of Computing*

O'Shaughnessy, D. (2012). Optimal Exchange Betting Strategy for Win-Draw-Loss Markets. *Proceedings of the 11th Conference on Mathematics and Computers in Sport*.

# TEAM RUNS PREDICTION VARIABILITY IN T20 CRICKET


Jonathan Sargent

*Stats On, Melbourne, Australia*
jon@statson.com.au



## Abstract

This paper explores factors contributing to the variance in pre-match and in-play T20 cricket team runs projections. Pre-match runs projections for Team *A* ($R_A$) and Team *B* ($R_B$) are mostly dependent on venue and weather effects, and a predicted winning margin, estimated from the pre-match win probability for each team. As well as guiding pre-match betting on total runs markets, $R_A$ and $R_B$ act as weights for the in-play runs projections and match odds calculation, with decaying influence as the match progresses. Differences between observed and expected team runs totals, $Diff(R_i) = Obs(R_i) - Exp(R_i)$, for team *i*, seem to be mostly influenced by physical conditions, including the state of the pitch, and player performance. Of particular interest to this research was whether to revise $R_B$, given $R_A$ at the completion of the first innings; specifically, whether significant +/- $Diff(R_A)$ at the change of innings was the result of good/difficult batting conditions or above/below expected player performance. The former suggests $R_B$ should also be increased/decreased; the latter is more complicated and requires a statistical interrogation of *A*'s batting and the skill through *B*'s potential batting line-up and *A*'s bowling attack, prior to any $R_B$ adjustment. The paper then demonstrates the value in addressing $R_A$ and $R_B$ by comparing an in-play profit statement from the recent ICC T20 tournament, to a simulated one should the runs projections have been held constant.



*Keywords*: T20 cricket, team runs projection, in-play betting

# PLAYER RATINGS IN PGA GOLF

Bradley O'Bree [a,b], Ian Grundy [a]

[a] *School of Mathematical and Geospatial Sciences, RMIT University, Melbourne*
[b] *Corresponding author: bradley.obree@rmit.edu.au*

## Abstract

There are a small number of applied ratings systems currently in operation in PGA golf. The most well-known would have to be the Official World Golf Rankings (OWGR), which is an example of an accumulative player rating system. Each ratings system has been developed to measure one particular aspect of player performance, be it expected difference in scores, such as in Sagarin Ratings, or the accumulation of tournament success, such as in OWGR. This research had two aims when developing a ratings system; the first was to measure the variation in player performance between rounds, and the second to determine a recommended sample size for performance analysis. A moving average based rating system was developed using all tournaments from the 2012 through 2015 PGA Tour seasons. The system used a relative measure of performance taken to be the ratio of player round score to average field round score. Model fits were very encouraging with adjusted $R^2$ measures consistently above .98. An algebraic manipulation of ratings equations showed that roughly 94% of variation in performance could be explained using a long-term measure of player strength, which can be defined as predominantly comprising scores greater than five rounds previous. The remaining 6% was described using results from more recent rounds, no more previous than four rounds back. Varying the total number of samples showed that a balance of predictive accuracy and minimisation of included samples occurred when a total of 20 rounds were used. Accuracy is seen to increase with the inclusion of more samples but plateau once 20 have been included. Results from this research provide a foundation for longitudinal analysis of performance in golf tournaments.

*Keywords*: PGA, golf, player ratings, sample size

## 1. INTRODUCTION

In recent years player ratings modelling has gained recognition as a medium between statistical research and end users in the form of fantasy league competitions. Player ratings give us a way of comparing the ability of players across different eras that don't necessarily compete against each other directly.

(Stefani 2010) provides an overview of ratings systems in sport. He notes there are three forms ratings systems can take, which include subjective, accumulative, and adjustive. Subjective systems involve the subjective awarding of points from referees or judges, as is the case for most martial arts competitions. Accumulative systems use the accrual of points awarded based on performance over several competitions and provide rankings for players or teams based on their current cumulative totals. Adjustive systems modify ratings between each competition based on some adjustment factor and the difference in the observed and predicted performance for a player or team. In this way adjustive systems are favoured in sports where future performance is to be predicted because ratings can increase or decrease based on the form of the player or team. This work has been expanded to include a survey of systems used in several international competitions in (Stefani 2011) and further to evaluate the predictive performance of a range of examples of each of the three types of systems in (Stefani 2012).

Perhaps the most well recognised rating system in golf is the Official World Golf Rankings (OWGR), which is an example of an accumulative system. Tournaments from a range of eligible international level professional tours are allocated rating points that players can earn based on their finishing position in a tournament field. The points available are determined by the importance or prestige associated with the status of a tournament (so that flagship events for example are allocated relatively more points) and influenced by the strength of the playing field, which changes based on the composition of world ranked top 30 and top 200 players competing within any given field. Ratings are weighted averages of results from the previous two years of competition with greater weights assigned to the most recent 13 week period. The OWGR have been used in a number of studies as a means of controlling for player strength in predictive modelling, as in (Nevill, A. M., et al 1997), where it was utilised to show there was insufficient evidence to suggest that home ground advantage was present in US and European golf tournaments.

Research has also been conducted into validating the OWGR as a ranking system, an important assumption that should be valid when used in predictive modelling. (Broadie and Rendleman 2012) sought to determine if rankings were biased by comparing the likeness of rankings between similar players from different professional tours. They claim that similar players from different tours should be similarly ranked if the system is not influenced by any bias. They used two unbiased methods for measuring scores, the first was a standard fixed effects regression model and the second another ratings method called Sagarin Ratings. Sagarin Ratings uses a player's won-lost-tied record against all other players within a tournament field to estimate the difference in typical round scores between players. Results are a reflection of the previous 52 weeks of performance, adjusting for the difficulty of a tournament with regards to the tournament field and the difficulty of a player's individual schedule. Results indicated that the system was biased against PGA Tour players, penalising them on average 26-37 OWGR ranking positions compared to non-PGA Tour players.

Other predictive rating systems exist that are used to predict relative measures of score as opposed to predicting actual scores. (Minton 2012) developed a rating system using the Strokes Gained metric, popularised by Mark Broadie in publications such as his book *Every Stroke Counts*, which uses an assessment of the quality of each shot to rate and infer performance.

The aim of this work is to determine the influence of sample size on the effectiveness of a predictive modelling system. Ideally there would be a balance between too few and too many inputs when gauging player performance over time. This has been investigated previously by (McHale and Forrest 2005) using an ordered logit forecasting model. They used tournaments from the 2003 US PGA Tour and the then current OWGR rankings to predict scores. Their results indicated the six most recent performances are the best indicators of future performance, and attributed this result in part to the OWGR rankings heavily favouring more recent performances. In this work we take a slightly different approach to solving the problem and arrive at a similar conclusion.

We organise the research aims into three research questions:

1. Is the size of predictable variation in sequential performances independent of player strength?
2. Is the size of predictable variation in sequential performances independent of tournament round?
3. Is there an optimal number of samples to include when developing a ratings model for golf performance?

## 2. METHODS

The methods section is comprised of three subsections. The first defines the performance metric of interest, the second explains the formulation of ratings models, and the third details the data used for the analysis.

### 2.1 MEASURING PLAYER PERFORMANCE

Round score ratios, calculated as player round score to field average round score, were chosen as the performance metric for this analysis. Typically fixed effect regressions are used to account for external effects like tournament difficulty a playing field, as in (Connolly and Rendleman 2008). If we assume however that tournament effects are similar across tournament fields, taking the performance ratio provides the measure of relative performance that lends itself to ratings systems. It is generally assumed that round scores are normally distributed, as previous work in score simulation (O'Bree, Bedford and Schembri 2012, O'Bree and Bedford 2014), performance modelling (Zumerchick 2008), and handicap optimisation (Swartz 2009) have suggested. However when performing this analysis regressions with normalised round scores provided poor adjusted $R^2$ values so performance ratios were used instead. We define $\alpha_{i,j}$ to be the ratio of round score $x_i$ for player $i$ in the $j^{th}$ tournament round to the field average round score $\overline{x_j}$.

$$\alpha_{i,j} = \frac{x_{i,j}}{\overline{x_j}}$$

### 2.2 THE RATINGS MODEL

We define the ratings model to be the linear regression equation of the sum of the $m$ previous ratio round score moving averages.

$$\alpha_{i,t} = \beta_1 \alpha_{i,t-1} + \beta_2 \left( \frac{1}{2} \sum_{k=t-1}^{t-2} \alpha_{i,k} \right) + \cdots + \beta_{m-1} \left( \frac{1}{m-1} \sum_{k=t-1}^{t-(m-1)} \alpha_{i,k} \right) + \beta_m \left( \frac{1}{m} \sum_{k=t-1}^{t-m} \alpha_{i,k} \right) + \varepsilon$$

The model equation is found using forward selection regression to keep the final model as simple as possible. We can manipulate the final equation knowing the average ratio score is equal parts of all included ratio scores to express the predicted score as a weighted average of previous scores without the any sort of moving average. For example, assume a final model equation that comprises the two and three point moving averages:

$$\alpha_{i,t} = 0.6\left(\frac{1}{2}\sum_{k=t-1}^{t-2}\alpha_{i,k}\right) + 0.4\left(\frac{1}{3}\sum_{k=t-1}^{t-3}\alpha_{i,k}\right)$$

The equation can be transformed to be represented as a function of only the previous performances, not moving averages of previous performances:

$$\alpha_{i,t} = \left(0.6\left(\frac{1}{2}\right) + 0.4\left(\frac{1}{3}\right)\right)\alpha_{i,t-1} + \left(0.6\left(\frac{1}{2}\right) + 0.4\left(\frac{1}{3}\right)\right)\alpha_{i,t-2} + \left(0.4\left(\frac{1}{3}\right)\right)\alpha_{i,t-3}$$

$$\alpha_{i,t} = 0.43\alpha_{i,t-1} + 0.43\alpha_{i,t-2} + 0.13\alpha_{i,t-3}$$

In this way, we can assess the change in final equation when increasing the size of *m*, allowing us to gauge the relative influence of previous scores as we include more and more previous samples.

### 2.3 DATA

Data for this research was sourced from publicly published scorecards from PGATour.com. The database contains all round scores from all standard medal-play tournaments from the 2012 through 2015 US PGA Tour seasons. A total of 69,599 scores were included in the analysis, with 170 incomplete rounds excluded. Standard tournament information accompanies the scores, including tournament name and round, as well as information regarding whether the tournament had a cut.

## 3. RESULTS

The results have been divided into three subsections. The first looks at regression results across all data. The second looks at regression results controlling for a simple measure of current player ability. The third looks at the results when controlling for the number of rounds remaining in the tournament.

There was a consistent trend that existed in final model equations as progressively more samples were included. When more than four previous performances ($m$>4) were included in a regression, the resulting final equation always took the form

$$\alpha_{i,t} = \beta_1\left(\frac{1}{4}\sum_{k=t-1}^{t-4}\alpha_{i,k}\right) + \beta_2\left(\frac{1}{m}\sum_{k=t-1}^{t-m}\alpha_{i,k}\right)$$

Where a total of *m* previous performances were made available in the regression. The first term included was always the *m* point moving average, and was only ever followed by the four point moving average. The fit of the regression model was never improved with the inclusion of the four point moving average but was always included due to statistical significance. Models fits were consistently exceeding .98 adjusted $R^2$ values.

### 3.1 THE GENERAL RESULT

The majority of the variation in current performance was explained by the long term moving average, as shown in Figure 1. The component was found to explain at least 94% of variation in current performances, with the proportion decreasing as more samples were included in regression.

Figure 2 shows the influence the four most recent performances have on current performance. The curves indicate that when eight or less previous performances are included, most of the variation in current performance can be explained using the four most previous performances *t-1* through *t-4*. When nine or more are used, most of the variation is explained using the less recent performances (*t-5* and older). This result suggests that when short of previous performances (less than nine), or when there are interruptions in a player's schedule, the best guess is based primarily on the last tournament played (or two tournaments if the player missed the cut).

### 3.2 CONTROLLING FOR PLAYER STRENGTH

The long term measure explains current performance best, but is this influenced by player ability?

Figure 3 shows what happens when we account for short term player strength. Players are divided into one of 10 equally sized groups based on their most previous round performance. Each group represents 10% of possible range of performance values, so that each group represents player strength in increments of 10% of all players. For example, say a player's previous performance saw him achieve a score ratio of 0.93, which for instance translates to the 23rd percentile of the score ratio distribution. This player would be allocated to group 3. Assessing the explained variation in current performances by these groupings shows the effects on model coefficients for the most recent four performances.

When the number of performances included was small (less than nine), the better players (groups 1, 2, etc.) had negative variance differentials, indicating less variation in current performance was explained by the last four performances when compared with the standard result across all players. The weaker players (groups 10, 9, etc.) mirrored this trend, with the previous four performances explaining more variation in current performance.

### 3.3 CONTROLLING FOR TOURNAMENT ROUND

The progression of a tournament could also see changes in the influence of previous performances. In the same way the effect of player strength was noted on the influence of the four previous performances, Figure 4 shows the effect of the tournament round expressed as the number of rounds remaining in the tournament. The graph shows that following the cut the influence of the previous four performances increases, while earlier rounds can be seen to have less influence. Interestingly, when more performances are included in the regression (moving averages span more performances) the effect of the tournament round decreases in size.

## 4. DISCUSSION

When determining exactly how to generate a ratings system the design relies heavily on the way the performance of the subject can vary over time. If performances vary greatly between competitions it can be difficult to find trends in individuals or teams that yield a means of measuring strength because of overwhelming noise in measurements. Conversely, if there is too little variation between competitors within a competition or between competitions, there is little use for ratings systems.

Results from this work have shown that when measuring relative performance of golfers over tournament competition that player performance is predictable using moving averages of previous performances represented as score ratios. Expectations for the current performance of a player are largely predicted using a long term moving average of these previous performances.

The base level model coefficients were found to vary amongst players of different strength and different levels of progression through a tournament. The stronger players tend to be more predictable using long term measures of their strength, while the weaker players tend to be more predictable when their most recent performances are analysed. This is likely due to the fact that better players are typically more consistent amongst multiple rounds and are successful because their skills are better developed and can handle the pressure of performing well across entire tournaments. The difference between coefficients dependent on player strength suggests that when developing ratings models the strength of the player needs to be considered.

The base level model coefficients were also found to vary throughout rounds within the same tournament. The influence of the previous four performances on model coefficients was found to increase as the tournament progressed. This does not come as a surprise. Typically tournaments have a cut after two rounds, meaning the better half of the playing field (for the current tournament) remain, so we can assume each player is playing well (or at least, better than average across the entire field). To be successful however you need to be one of the better players of the better half of the field, so your likelihood of a good performance is best determined by how well you've played in the previous rounds of the current tournament. As was the case with player strength, the recommendation here would be to vary any components of a ratings model that change ratings by the progression of a tournament, given the effect of tournament round is evident.

With regards to a recommended sample size, a rating system that includes 20 or more previous performances would see the benefits of the convergence of the proportion of variance explained by particular components; however there are certainly benefits to using smaller sample sizes (four or less) given player strength and tournament progression.

## 5. CONCLUSIONS

The results of this analysis have shown that the influence of previous performances on current predicted performance vary as a result of the number of performances included, the strength of the golfer, as well as the progression of the tournament. The structure of a ratings system in golf performance should therefore include these variables in its design.

**References**

Broadie, M., & Rendleman, R. J. (2012). Are the official world golf rankings biased?. *Journal of Quantitative Analysis in Sports, 9*(2), 127-140.

Brockwell, P. J., & Davis, R. A. (2002). An Introduction to Time Series and Forecasting. $2^{nd}$ edition. Springer Texts in Statistics. Springer.

Connolly, R. A., & Rendleman, R. J. (2008). Skill, Luck, and Streaky Play on the PGA Tour. *Journal of the American Statistical Association*, 103:481, 74-88.

McHale, I., & Forrest, D. (2005). The importance of recent scores in a forecasting model for professional golf tournaments. *IMA Journal of Management Mathematics*, *16*(2), 131-140.

Minton, R. (2012). Golf by the Numbers: How Stats, Math and Physics Affect Your Game. Baltimore, Maryland, The John Hopkins University Press.

Nevill, A. M., et al. (1997). Identifying home advantage in international tennis and golf tournaments. *Journal of Sports Sciences*, *15*, 437-443.

O'Bree, B., Bedford, A., & Schembri, A. (2012). An application of Bayesian Inference and Simulation in Golf. *Proceedings of the eleventh Australasian Conference on Mathematics and Computers in Sport*, A Bedford & A J Schembri (Eds), pp.194-199.

O'Bree, B., and Bedford, A. (2014). A Hole by Hole Simulation Model for Men's PGA Golf. *Proceedings of the eleventh Australasian Conference on Mathematics and Computers in Sport*, pp.76-82.

Official World Golf Ranking, accessed 1 January 2016, <http://www.owgr.com/about>

Sagarin, J. Men's Professional Rankings (PGA), accessed 1 January 2016, <http://rankings.golfweek.com/rankings/default.asp?T=mi>

Stefani, R. (2010). A World of Sports and Rating Systems. *$10^{th}$ Australasian Conference on Mathematics and Computers in Sport*, 1-12.

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

Stefani, R. (2012). Predictive Success of Official Sports Rating Systems in International Competition. *$11^{th}$ Australasian Conference on Mathematics and Computers in Sport*, 35-40.

Swartz, T. (2009). A New Handicapping System for Golf. *Journal of Quantitative Analysis in Sports*, 5(2), Article 9.

Zumerchik, J. (2008). Newton on the Tee: A Good Walk through the Science of Golf. Simon and Schuster, pp. 197-206.

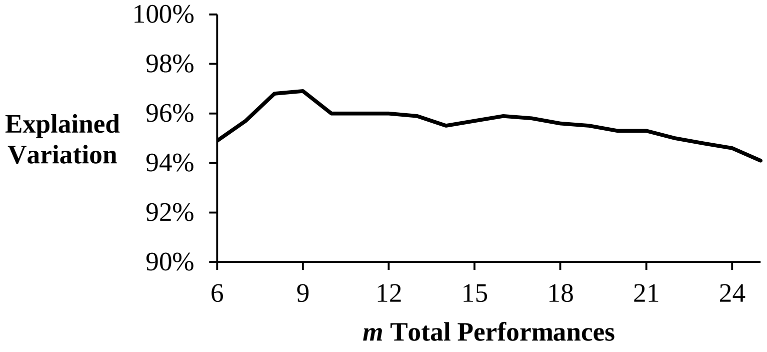



Figure 1. Explained variation in performance ***t*** from the moving average ***MA$_m$*** of each model comprising ***m*** total performances

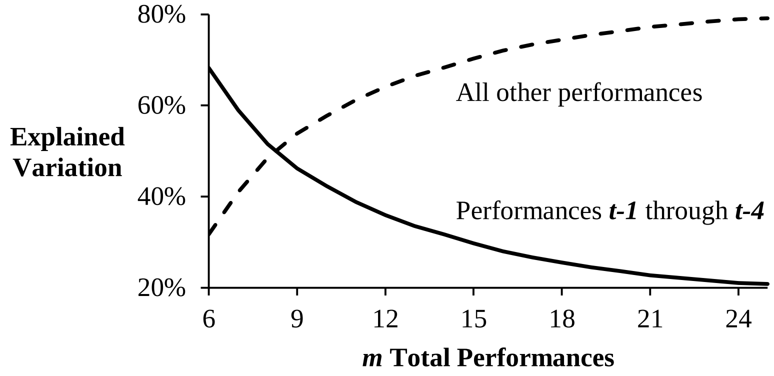



Figure 2. Explained variation in performance ***t*** using performances ***t-1*** through ***t-4***

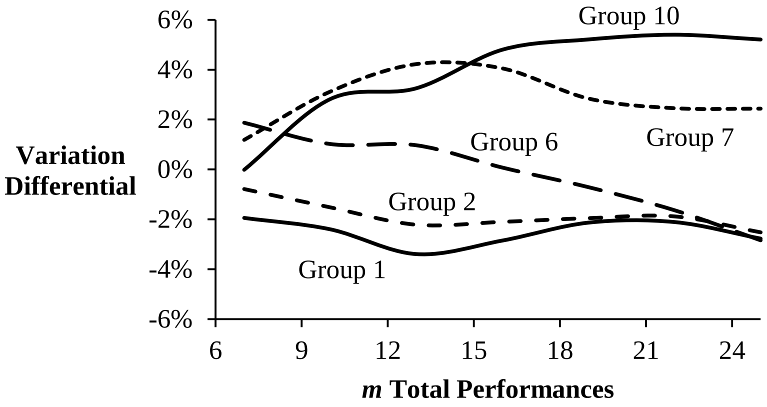



Figure 3. The effect of player strength on explained variation in performance ***t*** using performances ***t-1*** through ***t-4***

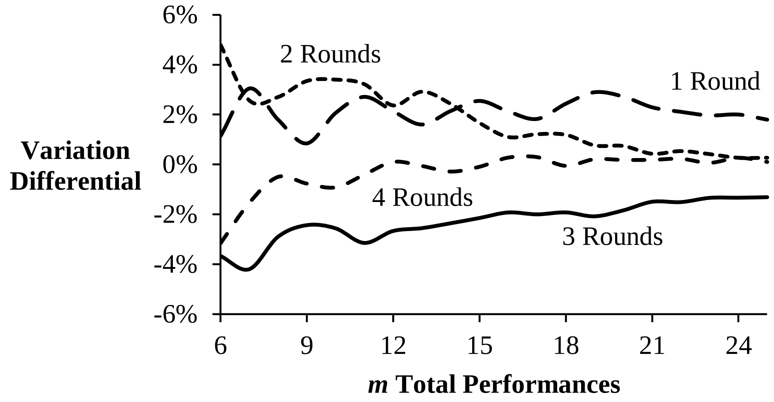



Figure 4. The effect of tournament round (rounds remaining) on explained variation in performance ***t*** using performances ***t-1*** through ***t-4***

# THE INFLUENCE OF PLAYER ABILITY AND POSITION ON TOURNAMENT SUCCESS IN PGA GOLF

Bradley O'Bree [a,b], Ian Grundy [a]
[a]*RMIT University, Bundoora*
[b] *Corresponding author: bradley.obree@rmit.edu.au*

## Abstract

Conventional wisdom on success in tournament golf tends to suggest an element of luck is required in most winning campaigns. For such a view to exist there needs to be a degree of uncertainty in current and final field position for players in the hunt to finish on top after four rounds of tournament competition. There a numerous questions one could answer in an effort to determine firstly if this is true, and if so, why it is true. A rather simple question though could be: what is more important, your position within the field or your ability as a player? The aim of this research was to determine the relative importance of position and ability-based variables when predicting tournament success on the US PGA Tour. A range of variables related to changes in position within a tournament field as well as simple player rating models were included in an initial linear regression to predict final tournament rank. The initial regressions were run for all tournaments from the 2012 through 2015 PGA Tour seasons and showed adjusted $R^2$ values ranging from 0.75 to 0.99 in final tournament ranks. While the predictive accuracy of the simple regression was promising, there remained the seemingly obvious issue of correlated predictors. As a result Principal Component Analysis (PCA) was used to condense the large number of correlated predictors into a smaller number of uncorrelated variables. PCA results reduced the large set of predictors into two components and accounted for approximately 85% of total variation in final tournament rank. The first component accounted for around 60% of predictable variation and comprised primarily of position-based variables. This result indicates position is in fact more influential on tournament success than player ability, though for this result to be concluded further analysis is required. This research helps to establish a means of simplifying predictive models in golf.

**Keywords**: PGA, golf, prediction, tournament success, principal component analysis

## 1. INTRODUCTION

It can be thought in general that a successful professional golfer is someone who not only plays well during tournaments but occasionally wins them. There are some tournaments that are much more acclaimed than others – for example, the four major tournaments in the US PGA Tour. Success has been seen by some as a purely money driven cause, where the player who earns the most has been the most successful. Usually though the success of a player relates to how many professional tournaments they've won.

Conventional wisdom on success in tournament golf tends to suggest an element of luck is required to beat an entire tournament field. For such a view to exist there needs to be a degree of uncertainty in current and final field position for players in the hunt to finish on top after four rounds.

A substantial amount of performance modelling in golf can be seen to focus on creating standards in performance statistics that correlate well with success. A number of studies have focussed on determining the more important performance statistic between driving distance and driving accuracy (having your ball land on the fairway once driven from the tee), as in (Hellstrong, Nilsson, & Isberg 2014). Ketzscher & Ringrose (2002) studied the use of summary statistics collected by the Professional Golf Association European Tour for predicting player success and concluded that golf was too complex for this to be achieved. Similarly, James (2007) concluded that the current performance indicators needed to be improved upon.

The research into these performance statistics provides interesting findings with regards to how statistics relate to scoring, but they don't provide the means for actually predicting success in terms of ranking as highly as possible. It isn't always the best player for the day in terms of putting who wins the tournament, or the player who hits the furthest average drive over the span of the tournament. Little of the research is concerned with using a player's position within a tournament field to predict success.

This work builds off of previous work by O'Bree & Bedford (2015) into using a player's tournament rank to predict player success at the end of a tournament. In this previous work, the profitability of wagering on outright tournament winners was examined using position-based variables to create final ranking probability distributions. The problem with these position-based models was that they could only be used independently; their predictive power couldn't be combined to enhance the predictability of final tournament outcomes. This was the result of the position-based variables being strongly correlated; they measure slightly different things but are essentially the same variables. Research into cricket performance by Manage & Scariano (2013) has addressed the issue of correlated predictors by utilising Principal Components Analysis (PCA). Their PCA facilitated the ranking and prediction of performances of both bowlers and batsmen in the 2012 Indian Premier League competition. The aim of this work is to assess the principal components of a range of position-based and ability-based variables to determine how much predictable variation in final tournament ranking can be retained while minimising the number predictive model inputs, all while accounting for the problem of the predictors being correlated.

## 2. METHODS

There are a number of ways to measure a player's position within a tournament field. To be able to best use predictive models we should have the best idea about the position of a player. We can use the distribution of scores within a field to gain insight into the context of a player's rank. The variables used in this analysis take three different forms:

- Field Rank: The rank of the round score for any individual round across all players, as well as the cumulative field rank of the player as a tournament progresses. For example, Adam Scott achieved the best round score in the second round and is currently ranked first in the field for the tournament.

- Shots Differential: The player's round score compared to course par score, compared to the field average round score, and compared to the field leader's score, for individual tournament rounds, and cumulative differentials as a tournament progresses. For example, Jordan Spieth scored a round of 69 on a par 72, while the leader achieved a score of 67 and the average field score was 73, giving his differential set of scores of {-3, -4, +2} respectively.

- Score Ratio: The ratio of the player's score to the course par, field average, and field leader's score, for individual tournament rounds and cumulative differentials as a tournament progresses. Taking Spieth as another example, his ratio score set would be $\{\frac{69}{72} = 0.96, \frac{69}{73} = 0.95, \frac{69}{67} = 1.03\}$.

Each variable is assessed at the conclusion of each day's play (between each round).

To estimate changing player strength across multiple tournaments, moving averages of each of these variables are taken across the previous five, 10 and 20 tournament rounds. The PCA will be conducted using the current position of a player and the player strength variables from one of the five, 10 or 20 round moving averages. These moving averages variables represent the measures of player strength.

Linear regression will be used to initially evaluate the predictability of final tournament rank using these variables, and accompanied with a correlational analysis of the variables with each other and with the final tournament rank.

All analysis will be conducted using all medal-play tournaments from the US PGA Tour seasons 2012 through 2015.

## 3. RESULTS

The results of the analysis will be divided into two subsections. The first will show the basic results from the linear regression analysis of the position and ability variables. The second will show the effect of the data reduction procedure through the PCA results.

### 3.1 LINEAR REGRESSION ANALYSIS

It comes as no surprise that the position variables were found to have reasonably strong correlations with final tournament rank. Figure 1 displays the strength of correlations by tournament round. It is clear that in each analysis the strong correlations are present between the measurements of position relative to the field average with final tournament rank, and that in each case correlations predictably become stronger as the tournament progresses. Figure 2 displays the correlations between the player-ability (moving average) variables from the last round of standard four round tournaments and final tournament rank. The same trend in the strength of correlations between variables exists as in Figure 1, and we observe that as moving averages include more data points the strength of the correlations weaken.

The linear regression analysis provided positive results in terms of predicting the final tournament rank of a player. Table 1 displays the adjusted $R^2$ model fits for each regression. In each case most of the predictors were statistically significant at the .05 level and model fits were very promising ranging from 0.75 for regressions using positions following the first round of the tournament to 0.99 for regressions conducted using round four positions. Clearly, multicollinearity is an issue in this analysis, as evidenced by the strong correlations between predictors in Table 2.

### 3.1 PRINCIPAL COMPONENTS ANALYSIS

The principal components analysis provided good insight into what influences tournament success, as demonstrated in Table 3. In general, the first principal component explained roughly 60% of variation amongst the variables. Adding the second increased this proportion to around 85%, and in each instance over 90% of variation is accounted for by the first three components. Interestingly, the benefit of greater progression through the tournament (the current round being later) only improved the performance of the first component, boosting the proportion of explained variation in final rank by between 9% and 16%. There was virtually no improvement to the second or third components, meaning the first two components account for any explanatory power that is to be gained as the tournament progresses. Similarly, when the first two components are paired together, explained variation in final rank across the different player-ability models (the three moving average models) was essentially the same. These results suggest that extracting the first two components provides enough explanatory power that the tournament round and player ability have no effect on model selection.

In an attempt to determine the answer to the main research question, Table 4 summarises the model coefficients for the first two components for both position and player-ability variables. We observe that most coefficients are positive, and that coefficients are much smaller for player-ability variables than for position variables, particularly in the 10 and 20 point moving average models. These coefficients are very close to 0 in essentially all variables for the two models with the greater number of samples included, which suggests that the scores determined by the principal components are virtually independent of player ability. Clearly, this indicates that position is more important to tournament success than player ability.

## 4. DISCUSSION

The aim of this work was to further develop a current rank-based approach to predicting tournament success. Initially, the approach involved using foundational models to evaluate the profitability of a wagering strategy in professional golf that utilised outright winner market prices. The main issue with these models was that they could not be used concurrently due to multicollinearity in model predictors.

A correlational analysis of these variables showed they provide reasonably strong insight into how successful a player is likely to be in the form of better final tournament rank. A principal component analysis of the variables themselves showed that a substantial proportion of variation in the set of variables could be summarised using only two or three components. An analysis of the component structures showed a strong dependency on the position-based variables, indicating that position within a tournament field is more important to success than player ability.

These results have the potential to contest the current direction research into predictive modelling in golf is currently taking. In the majority of predictive research the player-specific performance statistics are used to provide insight into how well a player is performing within a given tournament field. As is the case with sport however, the best players and teams don’t always win. Mentioned frequently in literature is the element of luck with succeeding in golf, even for the most highly skilled player (Connolly & Rendleman 2008). So one could question the usefulness of considering random predictors (such as fairway accuracy) over using fixed predictors, such as a player’s rank or the number of shots they trail the leader by. The fixed predictors provide insight that remains standardised across different tournaments and different situations, something that you can more safely rely on. Even if greens in regulation (the proportion of holes where a player reaches the green in expected shots, equal to hole par less two putts) correlates strongly with scoring, it would need to correlate more strongly than the fixed position variables to be of use when determining the winner of the tournament ahead of time.

If there is benefit for using these performance statistics, it is to add context to the player’s performance. A tournament leader that has five players trailing by a single shot is in much greater danger of surrendering his lead than does a leader who is eight shots ahead of the second ranked player, regardless of their respective strokes gained or putts-per-hole statistics. Primarily, tournament success should be measured as a product of players being in the strong positions at the right time in the tournament. Certainly in the early stages of the tournament, we only really have access to measures of player ability to provide evidence to back any claims we make about who the eventual winner of the tournament will be, but once holes have been completed position becomes king.

## 5. CONCLUSIONS

Results from this work have shown that a player's position within a tournament field is a better predictor of tournament success than longitudinal estimates of their strength as a player.

### References


Bowerman, B., & O'Connell, R. (1990). Linear Statistical Models: An Applied Approach. Duxbury Classic Series, 2nd Edition, USA.

Bro, R., & Smilde, A. K. (2014). Principal Component Analysis: Tutorial Review. *Anal. Methods, 6*, 2812-2831.

Connolly, R. A., & Rendleman, R. J. (2008). Skill, Luck, and Streaky Play on the PGA Tour. *Journal of the American Statistical Association*. 103(481). 74-88.

Hellstrom, J., Nilsson, J., & Isberg, L. (2014). Drive for dough. PGA Tour Golfers' tee shot functional accuracy, distance and hole score. *Journal of Sports Science, 32(5)*, 462-469.

James, N. (2007). The Statistical Analysis of Golf Performance. *International Journal of Sports Science and Coaching, 2*, 231-249.

Jolliffe, I. (2002). *Principal Component Analysis*. Springer Series in Statistics. Springer-Verlag, New York.

Ketzscher, R., & Ringrose, T. (2002) Exploratory Analysis of European Professional Golf Association Statistics. *Journal of the Royal Statistical Society,51(2)*, 215-228.

Manage, A. B. W, & Scariano, S. M. (2013). An Introductory Application of Principal Components to Cricket Data. *Journal of Statistics Education*, *21*(3).

O'Bree, B, & Bedford, A. (2015). Can betting in PGA golf be profitable? An analysis of common wagering techniques applied to outright winner market prices. *5th International Conference on Mathematics in Sport*, 139-143.

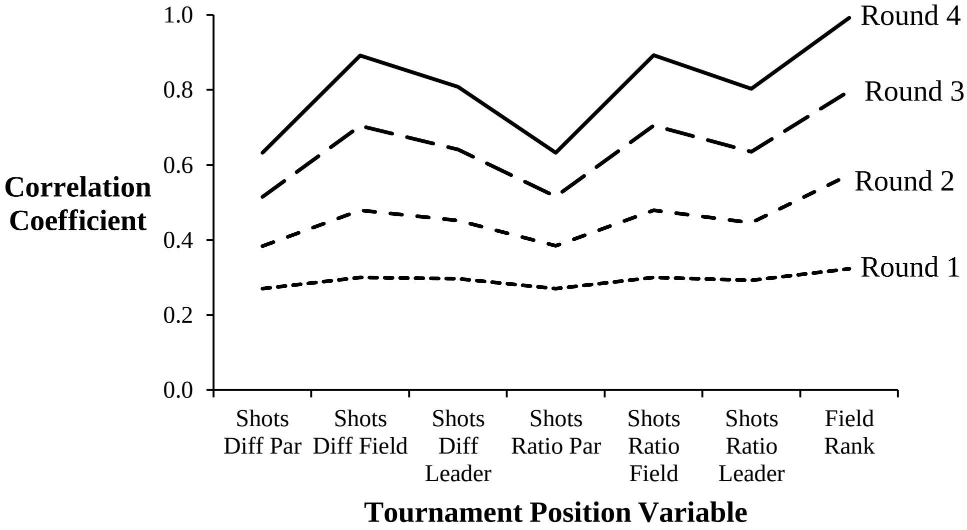



Figure 1. Strength of correlation between current tournament position variables and final tournament rank. All correlations were statistically significant at the .001 level.

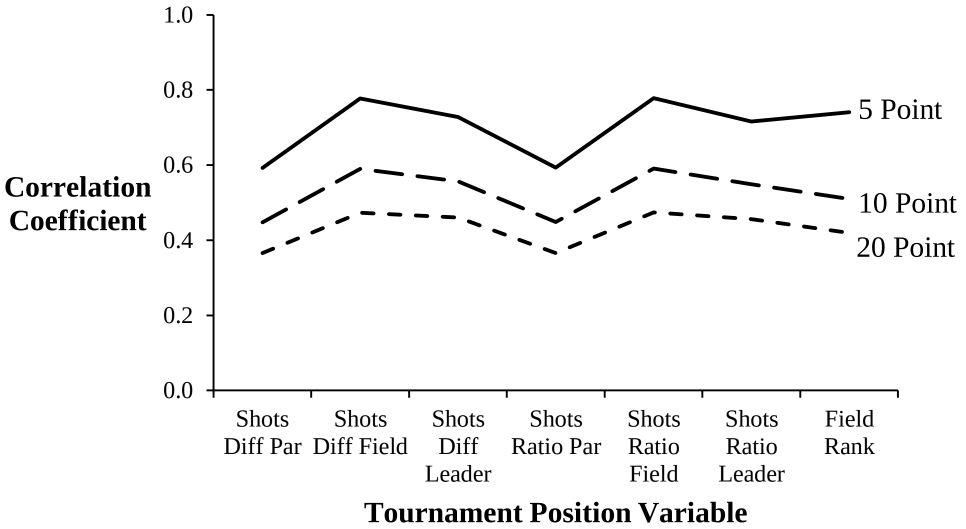



Figure 2. Strength of correlation between round four player-based moving average variables and final tournament rank. All correlations were statistically significant at the .001 level.

Linear Regression Model Adjusted $R^2$

| Tournament Round | 5 Point Moving Average | 10 Point Moving Average | 20 Point Moving Average |
|---|---|---|---|
| One | 0.75 | 0.76 | 0.76 |
| Two | 0.81 | 0.81 | 0.81 |
| Three | 0.89 | 0.90 | 0.90 |
| Four | 0.99 | 0.99 | 0.99 |

Table 1. Standard linear regression adjusted $R^2$ values

Position Variable Correlations

| | Shots Diff Par | Shots Diff Field | Shots Diff Leader | Shots Ratio Par | Shots Ratio Field | Shots Ratio Leader | Field Rank |
|---|---|---|---|---|---|---|---|
| Shots Diff Par | | | | | | | |
| Shots Diff Field | 0.68 | | | | | | |
| Shots Diff Leader | 0.54 | 0.83 | | | | | |
| Shots Ratio Par | 0.86 | 0.68 | 0.61 | | | | |
| Shots Ratio Field | 0.54 | 0.87 | 0.77 | 0.76 | | | |
| Shots Ratio Leader | 0.56 | 0.55 | 0.43 | 0.58 | 0.54 | | |
| Field Rank | 0.58 | 0.77 | 0.63 | 0.67 | 0.77 | 0.69 | |

Table 2. Correlation matrix for position variables. All correlations are significant at the .0001 level.

Cumulative Proportion of Explained Variation
in Final Tournament Rank by Principal Components

| Tournament Round | 5 Point Moving Average | | | 10 Point Moving Average | | | 20 Point Moving Average | | |
|---|---|---|---|---|---|---|---|---|---|
| | PCA 1 | PCA 2 | PCA 3 | PCA 1 | PCA 2 | PCA 3 | PCA 1 | PCA 2 | PCA 3 |
| One | 0.606 | 0.871 | 0.921 | 0.559 | 0.867 | 0.920 | 0.529 | 0.871 | 0.920 |
| Two | 0.627 | 0.831 | 0.910 | 0.569 | 0.832 | 0.907 | 0.535 | 0.837 | 0.905 |
| Three | 0.701 | 0.827 | 0.915 | 0.622 | 0.833 | 0.913 | 0.579 | 0.841 | 0.912 |
| Four | 0.765 | 0.868 | 0.930 | 0.674 | 0.836 | 0.926 | 0.619 | 0.847 | 0.925 |
| Average | 0.675 | 0.849 | 0.919 | 0.606 | 0.842 | 0.917 | 0.566 | 0.849 | 0.916 |

Table 3. Cumulative explained variation in final tournament rank by tournament round

Combined PCA 1 and PCA 2 Coefficients

| Position Variable | 5 Point Moving Average | | 10 Point Moving Average | | 20 Point Moving Average | |
|---|---|---|---|---|---|---|
| | Position | Ability | Position | Ability | Position | Ability |
| Shots Differential to Par | 0.71 | 0.65 | 0.52 | 0.06 | 0.51 | 0.04 |
| Shots Differential to Field | 0.15 | 0.11 | 0.54 | 0.02 | 0.55 | 0.01 |
| Shots Differential to Leader | 0.17 | 0.13 | 0.56 | -0.01 | 0.56 | 0.01 |
| Shots Ratio to Par | 0.71 | 0.65 | 0.52 | 0.06 | 0.52 | 0.04 |
| Shots Ratio to Field | 0.15 | 0.11 | 0.54 | 0.02 | 0.55 | 0.01 |
| Shots Ratio to Leader | 0.14 | 0.08 | 0.55 | -0.02 | 0.55 | 0.00 |
| Field Rank | 0.09 | -0.02 | 0.50 | -0.10 | 0.51 | -0.05 |

Table 4. Combined Principal Component coefficients by Position and Player-Ability variables

# USING A BROWNIAN MOTION TO CALCULATE THE IMPORTANCE OF POINTS IN BASKETBALL

Michael de Lorenzo [a, b], Ian Grundy [a]

[a] *School of Science, RMIT University, Melbourne, AUS*
[b] *Corresponding author: s3285802@student.rmit.edu.au*

## Abstract

Particular moments during a game of basketball can affect a team's probability of winning. Often, a player's performance can be subjectively scrutinised for their ability to perform during these critical moments. However, identifying the moments that are most critical to a team's chances of winning can be difficult. In this paper, we attempt to quantify the importance of points during a basketball game using a Brownian motion. Closely related to a random walk, and originally used to model the random motion of particles suspended in fluid, Brownian motion has been found in past research to provide a reasonable estimate of the win probability of a team during a basketball game, given the time remaining and the difference in score. By adjusting these probabilities and using an existing definition of point importance from tennis, it is possible to calculate the importance of particular points during a game of NBA basketball. Results from games across six seasons of NBA basketball suggest that the point importance can be reasonably quantified by using the Brownian motion. Furthermore, the point importance can be broken down into the importance of scoring the next one, two or three points. An exploration of these three components indicate that the three-point shot always has a greater importance than either a one-point shot or a two-point shot. We believe that the knowledge of point importance in basketball could be an effective tool in evaluating player performance.

**Keywords:** Point importance, basketball, Brownian motion, probability

## 1. INTRODUCTION

Particular moments during a game of basketball can have an effect on a team's probability of winning. These moments can occur early in the game if the match-up is lopsided or towards the conclusion of the game if the contest is close. Quantifying the importance of scoring the next point or points can help determine which moments are most critical during a basketball game. Knowledge of how important certain points are in a game of basketball could help coaches and team managers identify players who either perform well or struggle during key moments of a game.

In past research, point importance has been thoroughly explored in tennis, where it has been defined as the difference between two conditional probabilities: the probability that the server wins the current game given they win the next point, minus the probability that the server wins the current game given they lose the next point (Morris, 1977). González-Díaz, Gossner, and Rogers (2012) used this definition to assess tennis player's ability to perform when the stakes are high. It was found that there is a significant relationship between a tennis player's career success and their ability to win important points. The point importance definition by Morris (1977) has also been adapted into other sports such as badminton, where point importance has been explored in order to determine a serving strategy (Ladds & Bedford, 2010).

In basketball, there exists little research on measuring point importance. One paper by Goldman and Rao (2012) determined the importance of points during a game of NBA basketball using a differentiation of a normal distribution function in order to assess the effect that pressure may have on a team's free throw shooting and ability to retrieve offensive rebounds. It was found that in the last eight minutes the home team is significantly better at offensive rebounding, but is significantly worse at free throw shooting.

The aim of this paper is to measure the importance of points during an NBA basketball game using a Brownian motion. A continuous-time version of a random walk, and originally used to model the random movement of particles suspended in fluid, the Brownian motion has been found to provide reasonable estimates of a team's probability of winning during a basketball game, given both the time remaining and the lead of the home team (Stern, 1994). In this paper, we explore adjusting these probabilities into conditional probabilities in order to use the point importance definition conceived by Morris (1977).

The paper is broken down into the following sections: Section 2 details the methodology behind the Brownian motion model and the calculations required to generate the point importance; Section 3 provides an analysis of the point importance measure including graphical representation of the change during a game;

Section 4 provides a discussion about the results and details possible drawbacks; and finally Section 5 concludes the research and identifies possible amendments for future work.

## 2. METHODS

### BROWNIAN MOTION

Brownian motion is a mathematical model that is applied to understand the random movement of events, such as particles suspended in fluid and financial markets. Stern (1994) applied a Brownian motion model to basketball games to follow the progress of scores and generate win probabilities for the two competing teams. The author found that the Brownian motion model generates reasonable estimates of the home team win probability, despite some negative correlation appearing between some quarters.

The Brownian motion model for the progress sports scores, as described by Stern (1994), first requires the time to be transformed into the unit interval $t \in (0,1)$, which describes the proportion of the game that has expired. Let $X(t)$ represent the lead $l$ of the home team after time $t$, where $X(t)$ can be positive, negative or equal to zero. To apply the Brownian motion model, we must first assume that $X(t)$ can follow a Brownian motion model with drift $\mu$ (points per game advantage for the home team) and variance $\sigma^2$ per unit time (for further information regarding the assumption see Stern (1994)). Under the Brownian motion model, $X(t)$ can be modelled as:

$$X(t) \sim N(\mu t, \sigma^2 t) \tag{1}$$

Applying the random walk model, the probability that the home team wins, given they have a lead (or deficit) $l$ at time $t$, is calculated by:

$$\begin{aligned} P_{\mu,\sigma}(l,t) &= Pr(X(1) > 0 | X(t) = l) \\ &= Pr(X(1) - X(t) > -l) \\ &= \Phi\left(\frac{l + (1-t)\mu}{\sqrt{(1-t)\sigma^2}}\right), \end{aligned} \tag{2}$$

where $\Phi$ is the cdf of the standard normal distribution. Note that the win probability for the away team is found by taking the compliment of (2). In Stern (1994), the drift parameters $\mu$ and $\sigma$ were estimated using probit regression and were held constant for all teams, meaning that each game had the same pre-game advantage to the home team. This is a coarse assumption, however, as the strength of individual teams is ignored. This issue is raised in Stern (1994) with the author suggesting bookmaker line spread (money lines) may be a better substitute. Use of the bookmaker lines was explored by Glasson (2006) when applying a Brownian motion model to Australian Rules football, while pre-game estimates based on Elo ratings were applied by Ryall, Bedford and Glasson (2009).

For our model, we use a slight adjustment to the bookmaker's lines. The $\mu$ and $\sigma$ parameters were determined empirically using the observed margins for all of the matches from the previous season dependent on the bookmaker's line. This approach was used to account for the possibility that $\sigma$ may be dependent on $\mu$ or the bookmaker's line. A more appropriate method would be to break down the $\sigma$ by quarters through either a modified logarithmic function (Ryall et al., 2009) or through using bookmaker information to generate time-varying implied volatility (Polson & Stern, 2015).

### POINT IMPORTANCE

The point importance measure follows the definition conceived by Morris (1977) by calculating the difference between two conditional probabilities: the probability that a team wins the game at time $t$ given they score the next point, minus the probability that a team wins the game at time $t$ given they do not score the next point. Note here that the second probability equals (2) while the first probability is created by adjusting $l$ in (2). To adjust (2) into the required conditional probability, we must first consider that in basketball a single scoring possession for a team can result in one point (free throw), two points (two-point field goal), or three points (three-point field goal). Therefore, if we say that the potential change in $l$ is equal to $s \in \{1, 2, 3\}$, then the probability that the home team wins the match at time $t$ given they score $s$ points on their next possession is equal to the following:

$$P_{\mu,\sigma}(l,t,s) = \Phi\left(\frac{(l+s)+(1-t)\mu}{\sqrt{(1-t)\sigma^2}}\right) \quad (3)$$

The point importance is then defined as the difference between (3) and (2). For the away team, (3) is calculated by (-1)*$s$ to indicate that the lead $l$ of the home team is decreasing. By calculating (3) for both the home and away team, we are able to distinguish the point importance between the two teams, where scoring the next point may be more important to one team given the next score will give them the lead. This method of calculating point importance allows us to identify the impact that scoring the next $s$ points has on each team's win probability.

## 3. RESULTS

### MODEL PARAMETERS

Play-by-play data for 7,140 NBA games played over six seasons between 2006 and 2012 were collected from basketballvalue.com, while pre-game NBA bookmaker lines for the parameter $\mu$ were collected from bet365.com. The full play-by-play for all games were loaded into Microsoft Excel where the Brownian motion model was then applied. For games which ended in a tie (*X(1)* =0) and required overtime, the fractional time $t$ was re-set for the start of the additional five minute period. This was completed for the each additional overtime that was required to determine a victor.

As mentioned in the methods section, the parameters $\mu$ and $\sigma$ are determined by assessing all games from the previous season that had the same absolute pre-game money line as the current game of interest. A sample of the absolute money lines with average score difference and standard deviation for the 2010/2011 NBA season are presented in Table 1.

| ABS(Money line) | $\mu$ | $\sigma$ |
|---|---|---|
| 0 | -2.15 | 7.63 |
| 1 | 0.64 | 11.67 |
| 1.5 | 0.88 | 11.63 |
| 2 | -0.02 | 12.43 |
| 2.5 | 0.62 | 11.41 |
| 3 | -0.8 | 10.11 |
| 3.5 | 2.88 | 10.21 |
| 4 | 0.75 | 15.51 |
| 4.5 | 0.89 | 11.43 |

Table 1: 2010/2011 absolute money lines with average score difference ($\mu$) and standard deviation ($\sigma$)

### IN-PLAY PROGRESS

To assess how the point importance measured by a Brownian motion model progresses throughout a game, we explore two case studies. The first game we assess was played between New York (home) and Boston (away) on December 25th, 2011. The pre-game bookmaker line was equal to -5, with all games from the previous season with an absolute money line equal to 5 having an average score difference of -1.21 with a standard deviation of 12.30; resulting in a pre-game Brownian motion win probability of 53.91% for New York (46.09% for Boston). The lead for New York and the point importance for both teams throughout the contest are presented in Figure 1, with vertical lines distinguishing between quarters. Note that the point importance is broken down into the importance of scoring the next one, two or three points for both teams.

In Figure 1, it can be seen that the point importance for both teams responds to the lead of the home team, with the importance being high when the scores are close. This is most notable in the final quarter, where it is also observed that the point importance is higher for the trailing team. In fact, over the duration of the contest, it can be seen that the point importance for the trailing team is always higher than the leading team. This is a logical result, as scoring the next set of points in the game would be more critical to the trailing team as they must reduce the score margin in order to get back into the contest. It can also be seen that the three point importance for both teams is always higher than the one point and two point importance. Once again this makes logical sense, as a three-point shot rewards a team greater than a one-point free throw or a two-point shot.

With 16 seconds remaining in the contest, New York held a two point advantage. The point importance for all *s* scores at this stage is low for New York compared to Boston, where the two point importance and three point importance are quite high. In the dying stages of the contest, the three point importance is approximately double the two point importance. With six seconds remaining, Boston missed a three-point shot but collected the offensive rebound. They then missed a two-point shot as time expired, giving New York a two point victory. Overall, from this case study, it appears that the point importance measured using the Brownian motion has reasonably quantified how critical certain moments are during the contest for both teams. It has also been able to distinguish the importance according to which team is in the lead and which team is trailing.

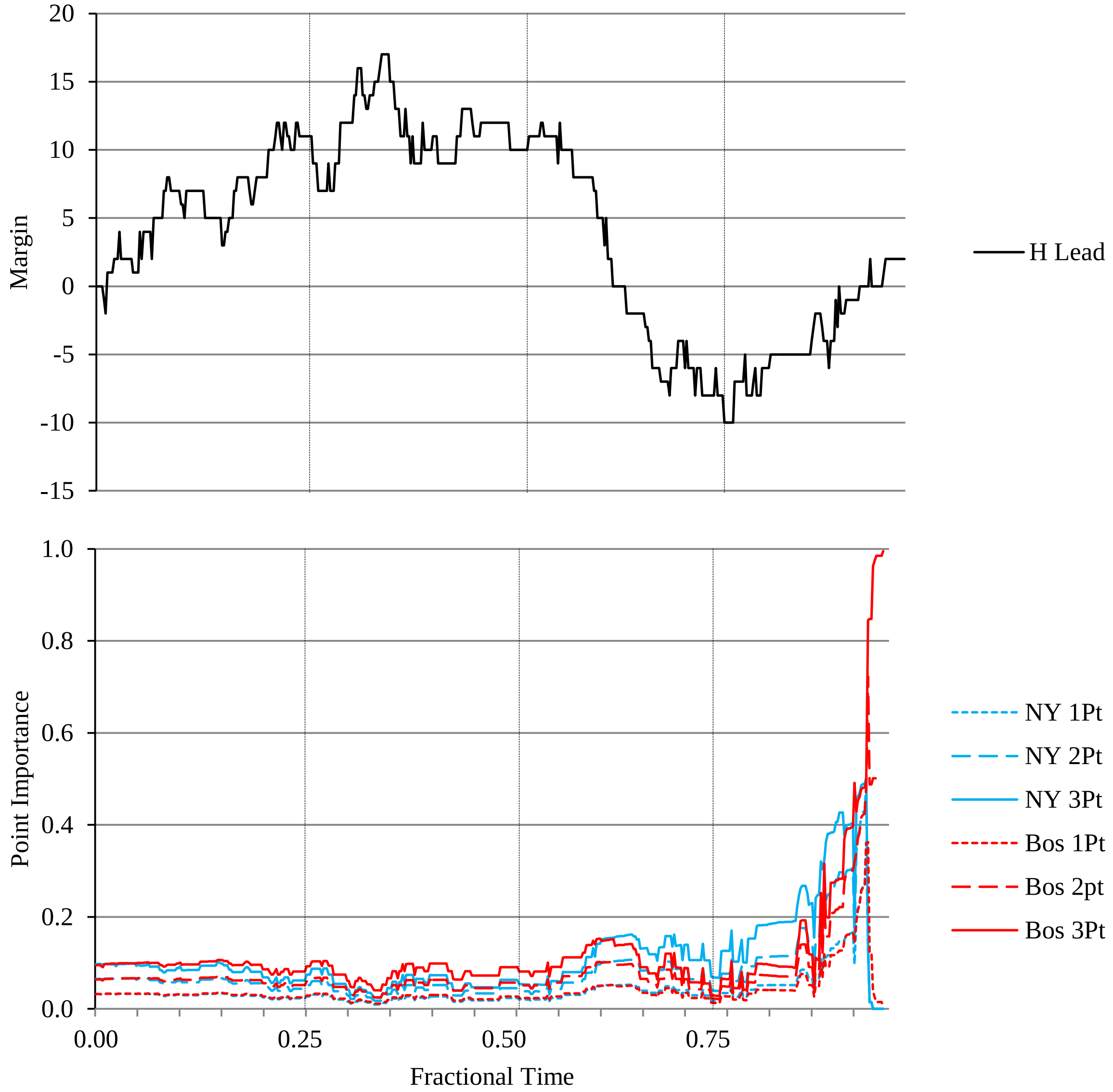



Figure 4: New York lead (top) and point importance for New York and Boston (bottom)

The second game we assess is a four-overtime contest played between Atlanta (home) and Utah (away) on March 25th, 2012. Evaluating this game allows us to explore how the point importance changes when the score is level at the end of regulation and overtime periods are required to determine a victor. The pre-game bookmaker line was equal to -1, with all games from the previous season with an absolute money line equal to 1 having an average score difference of 0.64 with a standard deviation of 11.67; resulting in a pre-game Brownian motion win probability of 54.66% for Atlanta (45.34% for Utah). Using the finding from the previous case study that the three point importance is always the highest, Figure 2 presents the three point importance throughout the contest, with vertical lines distinguishing between quarters and also the overtime periods.

As seen in Figure 2, the three point importance is low throughout the first half before increasing during the second. The increase in the final quarter is similar to Figure 2, indicating that the Brownian motion model is determining that points are most important in the final quarter. Once again this is a logical result, as the most critical moment during a contest would most likely occur during the final quarter when the score margin is small. Looking at the overtime periods, the point importance appears to increase quickly, which suggests that re-setting the fractional time at the start of the overtime period has worked correctly. Atlanta would eventually secure a four point lead with 13 seconds remaining in the fourth overtime. Utah would go on to miss their next two-point shot and lose the contest by six points.

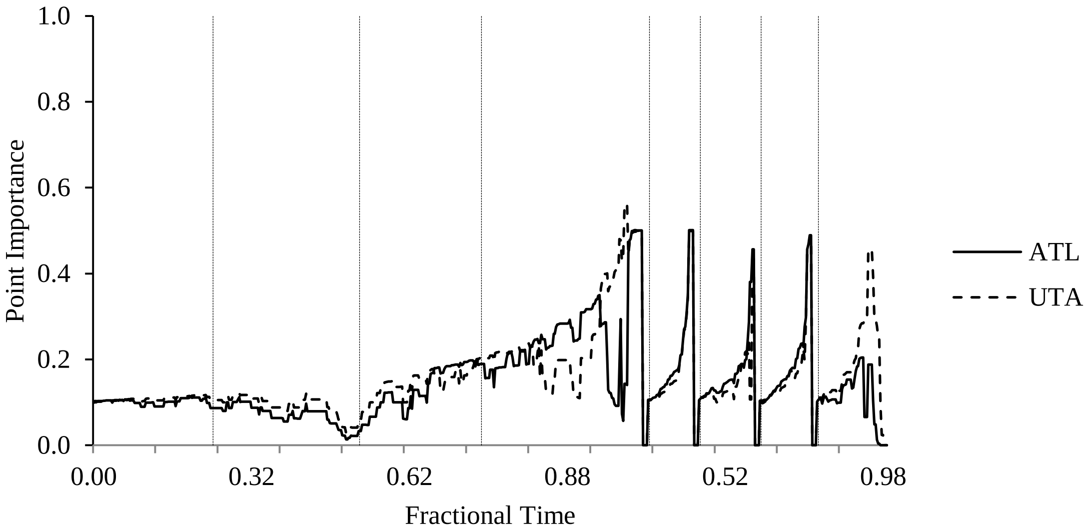



Figure 5: Point importance for Atlanta and Utah

## 4. DISCUSSION

The results from these case studies indicate that point importance can be reasonably quantified by applying a Brownian motion model. During a game, the maximum point importance occurs during the final quarter or overtime period, which is expected because this is when the most critical moments of a game would occur (especially if scores are close). This result indicates that the Brownian motion model can successfully quantify the importance of the critical moments correctly. This was observed in the Figure 1, with the point importance for the trailing team in both case studies being higher than the leading time.

An unsurprising observation from the first case study involved three-point shots always having a greater importance than a two-point shot or a one-point free throw. A three-point shot will always reward a team greater by increasing/decreasing the lead of the other team more than a two-point or one-point shot. Furthermore, the result also suggests that a trailing team should consider attempting more three-point shots as it always has a greater importance. This supports results on the risk of the three-point shot by Goldman and Rao (2013), who reported that teams should consider shooting more three-point shots when trailing during a contest.

While it has been discussed that the Brownian motion model has provided a reasonable method of calculating point importance during a basketball game there are drawbacks. Firstly, as discussed in Ryall et al. (2009), Brownian motion does not take into account the discrete nature of scoring, with the team that is ahead late often being predicted as the outright winner. Secondly, the model parameters that were implemented in this paper are not optimised and are calculated in a somewhat crude manor. Further refinement of these parameters could yield more accurate results, which may be achieved through standardising the pre-game money line and adjusting the parameters after each quarter (Polson & Stern, 2015; Ryall et al., 2009). The last drawback of calculating point importance with a Brownian motion model is that there is no information about which team is in possession of the ball at time *t*. Future work on distinguishing which team has possession of the ball could yield valuable benefits, as clearly point importance would be higher for a team is that is in possession of the ball with little time remaining, compared to if they did not have possession.

Despite the drawbacks, it is believed that calculating point importance using the Brownian motion model yields positive results that can be developed further in future research. It is also possible that the importance of points during a game of basketball can be used to assess player performance. This can be completed through isolating the most important moments during a game and collecting a player's performance statistics during these moments. It could be determined that a certain player shoots the ball with greater accuracy during important moments and is therefore a 'clutch' player, or perhaps some players can be found to turn the ball over more frequently during critical moments.

## 5. CONCLUSIONS

Throughout a game of basketball there are moments which can affect a team's probability of winning. In this paper, we have applied a Brownian motion model to calculate the importance of scoring the next points during a game of basketball in an attempt to identify critical moments. Results indicate that the Brownian motion model reasonably quantifies the importance of scoring the next set of points during a basketball game, with further work required on optimising model parameters and exploring player performance analysis.


### Acknowledgements

We would like to thank members of the RMIT Sports Statistics Research Group for their helpful suggestions and feedback.


### References


Glasson, S. (2006). A Brownian motion model for the progress of Australian Rules football scores. Paper presented at the Eighth Australasian Conference on Mathematics and Computers in Sport, Coolangatta, Australia.

Goldman, M., & Rao, J. M. (2012). Effort vs. concentration: The asymmetric impact of pressure on nba performance. Paper presented at the Sixth Annual MIT Sloan Sports Analytics Conference.

Goldman, M., & Rao, J. M. (2013). Live by the Three, Die by the Three? The Price of Risk in the NBA. Paper presented at the Seventh Annual MIT Sloan Sports Analytics Conference, Boston, USA.

González-Díaz, J., Gossner, O., & Rogers, B. W. (2012). Performing best when it matters most: Evidence from professional tennis. *Journal of Economic Behavior & Organization*, 84(3), 767-781. doi:http://dx.doi.org/10.1016/j.jebo.2012.09.021

Ladds, M., & Bedford, A. (2010). Using importance to determine a service strategy in badminton. Paper presented at the Tenth Australasian Conference on Mathematics and Computers in Sport (10 M&CS), Australia.

Morris, C. (1977). The most important point in tennis. In S. P. Ladany & R. E. Machol (Eds.), Optimal Strategies in Sport (pp. 131-140). Amsterdam: North-Holland Publishing Company.

Polson, N., & Stern, H. (2015). The implied volatility of a sports game. *Journal of Quantitative Analysis in Sports*, 11(3), 145-152. doi:10.1515/jqas-2014-0095

Ryall, R., Bedford, A., & Glasson, S. (2009). Exploiting in-the-run betting markets in Australian Rules football using a Brownian motion model. Unpublished work.

Stern, H. S. (1994). A Brownian Motion Model for the Progress of Sports Scores. *Journal of the American Statistical Association*, 89(427), 1128-1134. doi:10.2307/2290943

# ANCIENT OLYMPICS: EVENTS, SUPERSTARS, CHEATING, TECHNOLOGY AND WOMEN'S ROLE

Ray Stefani[a]

[a] *California State University Long Beach, USA*
[a] *Corresponding author: [Raystefani@aol.com](mailto:Raystefani@aol.com)*

## Abstract

We better understand the present and prepare for the future by understanding the past. We have analysed the list of winners from 776 BC to 277 AD, provided by the Persius Project. Events were added to maintain spectator interest, eventually totalling five sports (28 events): Athletics (6), Chariot Racing (9), Combat Sports (6), Equestrian Racing (3) and Artistic Performance (4).

Leonides of Rhodes was the Carl Lewis of antiquity, winning 12 championship wreaths, called ∑τεφανι, the Greek version of my family name. He won all three running events during four consecutive Games. The father of Alexander the Great, King Phillip II of Macedon, won three times. Nero won six times, albeit in events contrived for his benefit.

Athletes were disqualified for bribery, not for using performance enhancing drugs. Each cheater had to pay for part of a statue and have his name engraved. It took 14 statues: the first statue wasn't much of a deterrent.

The Games spawned technology. Clever systems were used to hold back runners for the start and to stagger the start of chariots to compensate for starting position. Modern research has shown that 5% can be added to the length of the (probably standing) long jump by properly swinging the weights that were carried. A leather cord and ring was used to provide leverage and spin stabilization for throwing the javelin.

Women were included in Greek sport. Unmarried women attended the Olympic Games. Married women were not to attend; but, one was Olympic champion, having owned and trained the winning chariot's horses. Unmarried women competed in the Heraia Games at Olympia. The running track was shortened to 500 Greek feet, 83% as far as men, exactly the ratio for female/male Olympic champion velocities in 1928, when women again competed in athletics. Female Olympic champions now run 90% as fast.

**Keywords**: Ancient Olympic Games, technology, starting mechanisms, women's equality, Heraia Games, athletics, long jump, halteres

## 1. INTRODUCTION

We better understand the present and prepare for the future by understanding the past. The Rio Olympics of 2016 are about to begin. The modern Games are now 120 years old, a relative short time span compared to their ancestral origin. The Olympic Games were first contested in 776 BC at Olympia, lasting more than 1000 years until abolished in 393 AD. During that time span, events were added to maintain the interest of spectators, as is done today. Efforts were made to entice support, for example from the Emperor Nero, much as we now seek support from television and commercial sponsors. Great athletes arose and became national heroes and Olympic icons. Organizers had to deal with and to punish cheaters, in adherence to their code of conduct. Technology was employed to improve performance and render fair starting conditions. Women were included in the sporting fabric at Olympia, based on the then religious practices.

Olympia was not the only site of sports competition during a four-year Olympiad. The Olympic Games would be contested on year one of a cycle, the Nemean and Isthmian Games on year 2, the Pythian Games on year 3 and then the Nemean and Isthmian Games would repeat on year 4. Collectively, these were called the Panhellenic Games or the Stephanitic Games. The latter term is used because each winner received a wreath called a ∑τεφανι in Greek, or Stefani, my family name.

We will now examine the events, superstars, technology and women's role in the Ancient Olympics. Because most of the reference titles are self explanatory as to material covered, most references will not be cited to avoid redundancy.

## 2. EVENTS

Thanks to the Perseus Project of Tufts University, USA, as sponsored by The Melon Foundation, The Annenberg Project and many others, ancient texts were scoured to create a list of 875 contested events from 776 BC through 277 AD. For each dated event, we have the winner and winner's home town. No winners were recorded after 277 AD. We placed that data base of 875 events into an Excel spread sheet. We deleted 42 events that had ambiguous event definitions or no event definitions. The remaining 833 events were sorted in various ways, to formulate as definitive a list of the contested events and winners as is currently possible.

| Event | Comments/Distance | First Year | Last Year | Times Held |
|---|---|---|---|---|
| Artistic Performance (4) | | | | |
| Herald Competition | Gap of 420 years until 65 AD | 396 BC | 261 AD | 12 |
| Lyre Playing | | 65 AD | 65 AD | 1 |
| Tragedy Competition | | 65 AD | 65 AD | 1 |
| Trumpeter Comp. | | 396 BC | 217 AD | 19 |
| Athletics (6) | | | | |
| Diaulos | X2, 384 m | 724 BC | 153 AD | 37 |
| Diaulos in Armor | X2, 384 m | 520 BC | 185 AD | 29 |
| Dolichos | X7-24, 1344-4608 m | 720 BC | 221 AD | 29 |
| Pentathlon | Discus, Javelin, Long Jump, Stadion, Wrestling | 708 BC | 241 AD | 31 |
| Stadion | X1, 192 m | 776 BC | 269 AD | 250 |
| Stadion-Boys | X1, 192 m | 632 BC | 133 AD | 30 |
| Chariot Racing (9) | | | | |
| Apene | 2 mules, x6, 7.2 km | 500 BC | 456 BC | 4 |
| Chariot Race | | 65 AD | 65 AD | 1 |
| For Foals | | 65 AD | 65 AD | 1 |
| 10 Horse Chariot | | 65 AD | 65 AD | 1 |
| Synoris | 2 horses, x6, 7.2 km | 408 BC | 60 AD | 14 |
| Synoris-Foals | 2 foals, x6, 7.2 km | 96 BC | 1 AD | 3 |
| Tethrippon | 4 horses, x12, 14.4 km | 680 BC | 241 AD | 59 |
| Tethrippon-Foals | 4 foals, x12, 14.4 km | 372 BC | 153 AD | 7 |
| Foals' Chariot Race | | 48 BC | 48 BC | 1 |
| Combat Sports (6) | | | | |
| Boxing | | 688 BC | 25 AD | 58 |
| Boxing-Boys | | 540 BC | 89 AD | 37 |
| Pankration | | 648 BC | 221 AD | 69 |
| Pankration-Boys | | 200 BC | 117 AD | 8 |
| Wrestling | | 708 BC | 213 AD | 66 |
| Wrestling-Boys | | 632 BC | 97 AD | 32 |
| Equestrian Racing (3) | | | | |
| Foal Racing | X6, 7.2 KM | 256 BC | 72 AD | 7 |
| Horse Racing | X6, 7.2 KM | 648 BC | 197 AD | 25 |
| Mare Racing | X6, 7.2 KM | 496 BC | 496 BC | 1 |

Table 1: Ancient Olympics by Sport (x6 means 6 lengths or circuits)

In Table 1, there were five sports (28 events) including artistic performance (4), athletics (6), chariot racing (9), combat sports (6) and equestrian racing (3). Among the artistic performance events, the herald and trumpeter competitions reflected a then-practical skill. Since there were no loud speakers, heralds and trumpeters were employed to announce and control public events as well as warfare. They had to be heard clearly over large distances. Oddly, two events were held only in 65 AD and the herald competition was held in 65 AD after a long hiatus.

In athletics, the most common running event was the stadion (giving us the word "stadium"), run over one length of the 600 Greek foot track, measured at 192 m. The other running events were multiples of the stadion. The pentathlon was an elimination competition including the discus, javelin and long jump (only contested in the pentathlon) plus the stadion and wrestling. The final competitors wrestled for the "Stefani".

The hippodrome at Olympia contained a 1.2 km course. Chariot races with two animals covered six circuits and with four animals, 12 circuits. Notice that three chariot races were contested only once, in 65 AD.

A race with two horses or foals was called a synoris while with four such animals; the race was called a tethrippon. Combat events include boxing, wrestling and the no-holds-barred pakration. The three equestrian events covered six circuits of the hippodrome, with competition for foals, horses and mares.

Of all 833 contested events, 49% were in athletics, 32% in combat events, 11% in chariot racing, 4% in equestrian racing and 4% in artistic performance. Clearly, the main interest was in athletics and combat competition, covering 81% of all contested events. It is interesting that artistic performances comprised 4% of events contested. The Second World Mind Sport Games (chess, bridge, checkers, go and Chinese checkers) were held just after the 2012 Olympics, with 29 events, 10% as many of the 300 physical sport events in the summer Games. Thus, the modern world, as the ancient world, honours mental and physical prowess

## 3. SUPERSTARS

The 833 event results were sorted by winner, to determine the top event winners of antiquity. Two famous names emerged, one of which explained anomalies in the events list. Nero won six events in 65 AD: the three chariot racing events and the two artistic performance events contested only in 65 AD and the herald competition in 65 BC that was held after a 420 year hiatus. There is an old adage that "Nero fiddled while Rome burned". Actually, Nero won the lyre playing competition. It was probably a bad career choice to defeat Nero, if anyone other then Nero actually competed in those six events. One can only wonder how many ten-horse chariots can race at a time. Organizers obviously wanted Nero's patronage. The other famous name was the father of Alexander the Great, King Philip II of Macedonia, who won the synoris (348 BC), the tethrippon (352 BC) and the horse race (356 BC). Defeating him was probably a bad career choice too.

Only three athletes won more events than Nero, although Nero's wins are highly questionable. In athletics, Leonides of Rhodes won 12 events: winning each of the stadion, diaulos and diaulos in armour for four consecutive Games starting in 164 BC, thus spanning a 12-year period. Given that all the youth of Greece wanted to win those same running events and that one slip in any one heat would have eliminated him, Leonides is arguably an Olympian for the ages. In our day, Carl Lewis won 10 Olympic gold medals in athletics from 1984 to 1996, including the 100 m run, 200 m run, long jump and 4x100 m relay. His four long jump wins covered a 12-year span, similar to Leonides' span. The second most prolific winner in the Ancient Olympics was Herodoros of Megara who won the trumpeter's competition nine consecutive times starting in 328 BC, thus spanning a remarkable 32 years. Third was Astylos of Kroton, who had been top winner in athletics before Leonides, having won seven times from 480 BC: winning in the stadion (three times), diaulos (once) and diaulos in armour (three times).

Four athletes won six events legitimately. Winners in wrestling were Hipposthenes of Sparta (from 632 BC) and Milon of Kroton (from 540 BC). Winners in athletics were Chionis of Sparta (from 664 BC) and Hermogenes of Xanthos (from 81 AD).

One of the most irrepressible athletes was Sostratos of Sikyon, nicknamed "Mr. Fingertips". He won the no-holds-barred pakration at Olympia three times (from 264 BC), the Isthmian and Nemean Games 12 times and the Pythian Games two times for a total of 17 wins. His winning record can be traced to his unusual ability of breaking his opponent's finger tips, thus his nickname

.

## 4. CHEATING

In the modern Olympics, there has been a concerted crackdown on performance-enhancing drugs since 1968 via drug testing. Such drugs were not illegal in the Ancient Games nor in the early years of the modern Games. In the ancient Games, athletes openly used opium juice, hallucinogens, strychnine and wine. As late as 1904, Thomas Hicks was given strychnine and brandy to help him win the marathon. The ancients were conscious of the need for proper nutrition (done legally as today). They ate large amounts of meat (providing protein), took herbs (which acted as supplements). They had a rudimentary idea about oxygen transfer and testosterone when they ate animal hearts and testicles.

Bribery was their form of cheating. As today, athletes pushed the boundaries of legality in search of monetary rewards. The organizers disqualified those who were caught bribing. Further, the guilty had to chip in for statues of Zeus with their names engraved thereupon. Was that a deterrent? If one statue was sufficient, the answer would be yes. In fact, over 1000 years, a total of 14 such statues, the Zanes, lined the athletes' path into the Olympic stadium. The lesson for today is that human nature has not changed: there will continue to be a prolonged battle between anti-doping authorities and the audacious rule-bending athlete.

## 5. TECHNOLOGY

Two parallel starting groves for the runners' toes are still in place at Olympia, with posts separating the 22 positions. In order to start those 22 runners fairly, a rope mechanism, the *hysplex* was created as in Figure 1 from Archaeology Archive (2015), taken at Stephen Miller's re-enactment at Nemea in 1993. See also Miller (2004) for a through discussion of Ancient Greek Athletics. In the left photo of Figure 1, starters in yellow at each end are cranking a narrow post from our left to our right, twisting a rope which would force the post to fall left were it not for restaining pins about to be put in place. The athletes are then restrained by two sets of connected ropes and the posts between them. In the centre photo, the head starter has yanked a rope, releasing the two restraining pins, causing the posts to fall to our left taking the ropes with them. The runners move to our left, where the outside runners have a small advantage, because the ropes fell from outside to inside.

The right most photo of Figure 1 shows the starting alignment for six pairs of chariots, which will move to our right after the start. The hippodrome was a flat 1.2 km oval with very sharp corners. During a race, the chariots nearest to the centre divider had to slow down on each turn so as not to impact outer chariots. The outermost chariots had to travel farthest around each curve. The chariots in the centre of the track could travel at constant speed. In the right photo of Figure 1, a dolphin-shaped lever controls the starting gates for each pair of chariots. The leftmost pairs start first. As they ride past the next pair, the gates open, giving the left most chariots an advantage. That pattern continues. When the right-most pair is released, the chariots form the mirror image of the starting alignment, compensating the outer chariots about to be disadvantaged by the turns.


Figure 1: Starting Mechanism for Running Races (Left and Center) and for Chariot Races (Right)

Figure 2: Technique for a Two Footed Long Jump with Weights

The long jump contestants in the pentathlon had to jump with halteres, smoothed weights, in each hand, weighing 1.5 to 5 kg. Minetti and Ardigoi (2002) found that trained athletes could gain 5.7% carrying a 2 kg weight with the optimum weight being 5-6 kg. Huang et al. (2005) established a gain of 4.5% using 2 kg weighs with an optimum weight equal to 8% of body mass. The conclusion is that a 5% gain is possible with proper technique. The best weight is 8% of body mass for both researchers.

An epigram indicated that Phayllos of Kroton once jumped 55 feet (16.3 m). He competed in the Pythian Games in 482 and 478 BC. Researchers from KU Leuven, The Ancient Long Jump and Phayllos (2012), indicated that after eight weeks of training, athletes jumped 15 m using five two-footed jumps, five for the number of events in the pentathlon. The left part of Figure 2 shows the resulting technique through the forward thrust. The right photo from KU Leuven shows the landing, as shown on a contemporary urn.

Another bit of technology was twisting a cord around the javelin, ending in a loop for the thrower's finger. That loop provided leverage and, as the cord unwrapped, the javelin was spin-stabilized, Miller (2004), 69-70.

## 6. WOMEN'S ROLE

This information is drawn primarily from Were Women Allowed at the Olympics (2016) and Miller (2014), pages 150-159. The religious practices of the day defined the women's role in ancient Greek sports, based on the gender of the god to whom a competition was dedicated. What we now call the Olympic Games was dedicated to Zeus. Being a male god, only men were allowed to compete. Unmarried women could and did attend. The High Priestess of Demeter was an honoured dignitary. Married women were not supposed to attend; however, Kyrniska of Sparta was a double Olympic champion in 396 and 392 BC, having owned and trained the winning chariot horses in the tethrippon. She accepted her laurel wreathes outside the stadium. Kallipateira of Rhodes trained her son inside the stadium but was discovered. She was pardoned since her father, three brothers a nephew and her son were champions. Thereafter, trainers had to be naked.

Women competed in the Heraia Games, so named because these games were dedicated to Zeus' mythological wife Hera, contested in a different year from the Olympic Games; but in the same Olympic stadium. Unmarred women competed. Married women served as officials and trainers. We do not know if men were allowed to attend. The Greek government empowered the so called Sixteen Women, all married, to coordinate female sports in all of Greece. Women competed in three age groups, over a stadion, reduced from 600 Greek feet to 500. This implies that women were assumed to run about 5/6 or 83% as fast as men. In fact, when women resumed Olympic athletics competition in 1928, the female champions ran 83% as fast as their male counterparts, Stefani (2014). Today, female Olympic champions in athletics run about 90% as fast as the male champions, Stefani (2014).

Current tradition calls for the Olympic flame to be lit at Hera's shrine in Olympia. That lighting provides a symbolic equality for women, given that men and women now both compete at the same time, in the same place and in nearly equal numbers of events.

## 7. CONCLUSIONS

Many of the features and trends of the Ancient Olympics exist today and are likely to continue into the future. As today, the most popular sport was athletics while sports and events were added to maintain spectator interest. Financial patronage was sought by including Nero and King Philip II of Macedonia as Olympians while we seek television and sponsor revenue. The greatest winner in athletics was Leonides who won 12 events over a 12-year period. Carl Lewis won 10 events over a similar 12-year period, apparently the competitive life span in athletics. Organizers sought to catch bribers over the entire span of the Ancient Games as evidenced by the need for 14 statues with cheaters' names inscribed. We are likely to continue the contest between anti-doping efforts and athletes seeking an illegal edge. Technology consisted of equipment to produce a fair start in running and chariot racing and technique for the long jump and javelin, much as we seek such technology today. The role of women was delimited by the gender of the god to whom competition was dedicated. The ancient and modern worlds are linked in that the torch for today's Games is lit at Hera's shrine while men and women now compete together, producing an equity for women only begun in Ancient Greece.

## References


Ancient Greeks Use Performance Enhancing Drugs. (2015, March 23). Retrieved from http://sportsanddrugs.procon.org/view.timeline.php?timelineID=000017.

Archaeology Archive. (2015, December 28). Starting Rope System for Running, Retrieved from http://archive.archaeology.org/online/features/olympics/stadia.html.

Foundation of the Hellenic World. (2015, October 24). Winners of the Ancient Olympics, Retrieved from http://www.fhw.gr/olympics/ancient/en/db.html.

From Ancient Olympics to Athens of 1896. (2015, October 24). Retrieved from http://www.fhw.gr/olympics/ancient/en/db.html .

Huang, C. et al. (2005). The Effect of Hand Held Weights on Standing Long-Jump Performance, *ISBS Conference, Beijing.*

Miller, S.G. (2004). *Ancient Greek Athletics*, New Haven, CT: New Haven.

Mineti, A.E. and Ardigo, L.P. (2002). Biomechanics: halteres used in ancient Olympic long jump. *Nature* 420, 141-142.

Olympic-Legacy 1 (2015, October 23). General information about the Ancient Olympics, Retrieved from http://olympic-legacy.com/.

Olympic-Legacy 2 (2015, October 24). Ancient Olympic Events, Retrieved from http://www.pe04.com/olympic/olympia/events_o.php.

Perseus Project 1 (2015, September), Ancient Olympic Events and General Information, http://perseus.tufts.edu/Olympics/sports.html and http://www.perseus.tufts.edu/Olympics/index.html.

Perseus Project. 2 (2015, September), Funding of the Perseus Project, Retrieved from http://www.perseus.tufts.edu/hopper/.

Starting Mechanism for Chariot Racing (2012). Retrieved December 2015from http://ancientolympics.arts.kuleuven.be/eng/TA012aEN.html .

Stefani, R.T (2014). Understanding the Velocity Ratio of Male and Female Olympic Champions in Running, Speed Skating, Rowing and Swimming, *Proceedings of ANZIAM Mathsport 2014, Darwin, Australia*.

The Ancient Long Jump and Phayllos (2012). Retrieved from http://ancientolympics.arts.kuleuven.be/eng/TC003EN.html on September 19, 2013.

Were Women Allowed at the Olympics? (2015, December). Retrieved from http://www.perseus.tufts.edu/Olympics/faq5.html .

# ANALYSIS OF GAME STATISTICS ON TEAMS AND PLAYERS IN JAPAN RUGBY TOP LEAGUE

Makoto Kiuchi [a, b], Nobuyoshi Hirotsu [a]

[a] *Juntendo University*
[b] *Corresponding author: makotos621115@yahoo.co.jp*

## Abstract
In Japan, the rugby union has been becoming a popular sport, especially after the Japan national rugby team made three wins in the Rugby World Cup 2015, together with the fact that Japan will hold Rugby World Cup 2019. The aim of the Japan national rugby team is to advance to the final round in Rugby World Cup 2019 as the host country. In Japan, not so many quantitative studies have been conducted, especially for Japan Rugby Top League (JRTL). Here, we try to evaluate the teams and the players in JRTL, and compare them with world level teams and players. As the result of this study, in terms of team evaluation, the defense of the teams in JRTL seems to be weaker than that of Six Nations Tournament. In terms of player evaluation, most of the passers and place kicker as tackler are Japanese players in JRTL, although few Japanese players are place kicker as game controller and competing possession players. Japanese players seem to be weak in tackle, game control and competing possession. Therefore, JRTL would need to foster Japanese tackler, game controller and competing possession player to strengthen the Japan national rugby team.

**Keywords**: japan rugby top league, rugby union, game statistics

## 1 · INTRODUCTION
Rugby Union is a contact sport played by 15 players. These 15 players are assigned 10 different positions in the game. These include: forward (FW), prop(PR), hooker(HO), lock(LO), flanker(FL), number 8(NO8), backs(BK), scrum-half (SH), stand-off(SO), center(CTB), wing(WTB) and full-back(FB).

The Japan national rugby team won against the South Africa national rugby team in the Rugby World Cup 2015. Japan will hold the Rugby World Cup in 2019. The target of Japan national rugby team is to advance to the final round as the host country of the 2019 Rugby World Cup. In Japan, there is a domestic rugby union competition called Japan Rugby Top League (JRTL). There are 16 teams in JRTL and the players who belong to Japan national rugby team are selected from this league. Therefore, if the level of JRTL becomes higher, the Japan national rugby team will be strong. However, JRTL have not been studied from the view of comparison with other international top-level league. Thus, in this study we try to evaluate the teams and the players in JRTL and compare them with world level teams and players.

In terms of team evaluation in rugby, Ortega et al (2009) show the differences of the game statistics of the winning teams and losing teams of Six Nations Tournament (SNT). We analyze the SNT in the way of Ortega et al., and compare between JRTL with SNT. In terms of player evaluation, in recent years, the number of JRTL players coming from nations such as New Zealand, Australia, and South Africa has been increasing. Thus, it is possible to compare between Japanese and foreign players in JPTL. Using the quantitative approach to the evaluation of teams and players, we would like to raise the problems that the Japan national rugby team to achieve the target.

## 2. METHODS
DATA

In this study, data were collected using a smart phone mobile application called "Japan Rugby Top League official application". In terms of team evaluation, we use the data of the 2015-2016 season. As there are 16 teams in the league, there were 80 games in the season (10 games for each team). Actually, 78 game data were used except for two games because of withdrawn. On the other hand, SNT data were collected from the site of “rbs6nitons.com”. There are 6 teams in SNT and 15 games were played in the 2016 season (5 games for each team). Actually, 14 games were used for this study because of withdrawn. Data analysis was conducted based on play items shown in Table1. There are 3 groups in these play items: points scored, way in which points were scored, phase of play, way teams obtained the ball and how the team used it; and Game development, all of which are tactical aspects of the game.

In terms of player evaluation, we use the data of the 2014-2015 season. There were 112 games in the season which was not influenced by rugby world cup 2015. As each team played 14 games in the season, total time was 1,120(=14×80) minutes. In this study, we analyzed 171 players who played more than 700 minutes, about two-thirds of the total time. Data analysis was conducted based on 19 play items shown in Table 2.

| Points scored | Phases of play | Game development |
|---|---|---|
| Points scored | % Time in possession | Line breaks |
| Tries Scored | 1st time | Takles made |
| Successful conversions | 2nd time | Takles miss |
| Successful penalty goals | % Time in opponents half | % Takles completed |
| Successful Drop goals | Scrum won | Turnovers won |
| | Scrum lost | Passes made |
| | Line-out won | Possessions kicked |
| | Line-out lost | |
| | Set-piece ball won | |

Table1. Items studied in team evaluation

| Point scored | Ball touch | Point made | Tackle |
|---|---|---|---|
| Try scored | Pass | Offroad | Tackle assist |
| Conversions | Kick | Tackle break | Jackal |
| Penalty goals | Contact | Support | Turnover won |
| | Penalty | Line break | Turnover lost |

Table2. Items studied in player evaluation

STATISTICAL ANALYSIS

In team evaluation, the Mann-Whitney U test was carried out for analyzing the difference between the winning teams and the losing teams. In player evaluation, the principal component analysis was used in order to summarize the data of the 19 items. In addition, the players were divided into clusters using the cluster analysis based on the k-means method. The SPSS 22.0 statistical program was used for this analysis.

## 3. RESULT & DISCUSSION

TEAM EVALUATION

The average values, standard deviations, and medians of the values of the play items with regard to the winning teams and the losing teams in both JRTL and SNT are shown in Table 3.

For the point scored, the difference in tries score and successful penalty goals can be seen in the comparison between JRTL and SNT. The tries score of JRTL is higher than that of SNT in both the winning teams and the losing teams. In addition, the penalty goal of JRTL is lower than that of SNT. From this result, JRTL defense is likely to be weaker than SNT. This is because if the opposing defense is strong, attacking team tends to choose penalty goal rather than try as a way of scoring.

For the phases of play, the difference in line-out lost is not seen in the comparison between the winning teams and the losing teams of JRTL, although the difference in line-out lost can be seen in SNT. The line-out lost of the losing teams in SNT occurred more frequently than that of the winning teams. From this result, the line-out lost in JRTL does not seem to make a major influence to winning.

For the game development, the difference in terms of tackles missed and possession kick can be seen in the comparison between JRTL and SNT. The tackle missed of JRTL is higher than that of SNT in both the winning teams and the losing teams. In addition, the possession kick of JRTL is lower than that of SNT in both the winning teams and losing teams. From the fact that tackles missed occur more frequently and possession kick occurs less we would infer that the defense of JRTL is weaker than that of SNT. This is because if the opposing defense is strong, attacking team often chooses the boll carry rather than the possession kick.

| Items | Winner M | Winner SD | Winner Median | Loser M | Loser SD | Loser Median | P Value |
|---|---|---|---|---|---|---|---|
| Variables related to point scored | | | | | | | |
| Points scored | 36.3 | 12.0 | 34.0 | 16.5 | 7.1 | 17.0 | 0.00 |
| Tries Scored | 4.8 | 2.0 | 4.0 | 2.2 | 1.0 | 2.0 | 0.00 |
| Successful conversions | 3.7 | 1.8 | 3.0 | 1.6 | 1.1 | 2.0 | 0.00 |
| Successful penalty goals | 1.6 | 1.3 | 1.0 | 0.8 | 0.9 | 1.0 | 0.00 |
| Successful Drop goals | 0.1 | 0.2 | 0.0 | 0.0 | 0.1 | 0.0 | 0.10 |
| Variables related to phases of play | | | | | | | |
| % Time in possession of full time | 50.8 | 7.5 | 51.5 | 49.2 | 7.5 | 48.5 | 0.19 |
| % Time in possession of 1st time | 52.8 | 9.7 | 52.5 | 47.2 | 9.7 | 47.5 | 0.00 |
| % Time in possession of 2nd time | 48.8 | 11.3 | 49.0 | 51.2 | 11.3 | 51.0 | 0.17 |
| % Time in opponents half | 51.2 | 9.5 | 50.0 | 48.7 | 9.7 | 49.5 | 0.16 |
| Scrum won | 6.9 | 2.7 | 7.0 | 6.5 | 2.3 | 6.0 | 0.34 |
| Scrum lost | 0.6 | 0.9 | 0.0 | 0.7 | 0.9 | 0.0 | 0.45 |
| Line-out won | 10.8 | 3.3 | 11.0 | 9.9 | 3.1 | 9.0 | 0.08 |
| Line-out lost | 2.1 | 1.7 | 2.0 | 2.4 | 1.7 | 2.0 | 0.12 |
| Set-piece ball won | 17.7 | 4.2 | 18.0 | 16.4 | 3.5 | 16.5 | 0.04 |
| Variables related to game development | | | | | | | |
| Line breaks | 4.9 | 3.3 | 4.0 | 3.2 | 2.1 | 3.0 | 0.00 |
| Takles made | 126.2 | 35.9 | 125.5 | 121.1 | 35.7 | 114.5 | 0.32 |
| Takles miss | 23.4 | 8.5 | 23.1 | 24.8 | 10.0 | 21.9 | 0.60 |
| % Takles completed | 81.4 | 4.6 | 82.0 | 79.5 | 5.3 | 80.0 | 0.04 |
| Turnovers won | 8.2 | 3.6 | 7.2 | 5.6 | 2.7 | 5.0 | 0.00 |
| Passes made | 130.5 | 36.2 | 125.0 | 139.6 | 42.0 | 140.0 | 0.10 |
| Possessions kicked | 21.2 | 6.3 | 20.0 | 17.5 | 6.1 | 16.0 | 0.00 |

(a) Japan Rugby Top League

| Items | Winner M | Winner SD | Winner Median | Loser M | Loser SD | Loser Median |
|---|---|---|---|---|---|---|
| Variables related to point scored | | | | | | |
| Points scored | 31.1 | 15.1 | 28.0 | 16.1 | 5.7 | 16.5 |
| Tries Scored | 3.5 | 2.5 | 3.0 | 1.4 | 1.0 | 2.0 |
| Successful conversions | 2.4 | 2.0 | 1.5 | 1.1 | 0.9 | 1.0 |
| Successful penalty goals | 2.9 | 1.5 | 3.0 | 2.1 | 1.7 | 2.0 |
| Successful Drop goals | 0.0 | 0.0 | 0.0 | 0.1 | 0.3 | 0.0 |
| Variables related to phases of play | | | | | | |
| % Time in possession of full time | 50.7 | 7.6 | 51.7 | 49.3 | 7.6 | 48.3 |
| % Time in possession of 1st time | 57.4 | 8.5 | 57.5 | 42.6 | 8.5 | 42.5 |
| % Time in possession of 2nd time | 44.0 | 14.7 | 46.0 | 56.0 | 14.7 | 54.0 |
| % Time in opponents half | 50.8 | 7.8 | 52.0 | 49.3 | 7.8 | 48.0 |
| Scrum won | 7.2 | 2.1 | 7.0 | 4.9 | 2.9 | 4.5 |
| Scrum lost | 0.8 | 0.9 | 0.5 | 0.6 | 0.6 | 1.0 |
| Line-out won | 10.0 | 2.8 | 9.0 | 11.0 | 3.2 | 10.0 |
| Line-out lost | 0.9 | 1.1 | 1.0 | 2.1 | 1.2 | 2.0 |
| Set-piece ball won | 17.2 | 3.7 | 16.5 | 15.8 | 3.7 | 15.0 |
| Variables related to game development | | | | | | |
| Line breaks | 7.9 | 5.4 | 6.0 | 5.7 | 2.5 | 6.0 |
| Takles made | 123.1 | 20.0 | 121.0 | 123.4 | 38.2 | 116.5 |
| Takles miss | 15.9 | 6.1 | 15.5 | 17.5 | 3.6 | 17.0 |
| % Takles completed | 88.6 | 3.9 | 88.0 | 87.0 | 3.4 | 87.5 |
| Turnovers won | 7.7 | 4.2 | 7.5 | 5.6 | 2.3 | 6.0 |
| Passes made | 141.4 | 46.2 | 126.5 | 142.7 | 30.6 | 138.0 |
| Possessions kicked | 32.6 | 7.6 | 33.0 | 27.6 | 8.2 | 27.0 |

(b)Six Nations Tournament

Table 3. Differences between the winning teams and the losing teams as the game statistics

## PLAYER EVALUATION

Principal component analysis was conducted for player evaluation in JRTL using the 19 items shown in Table 2. Table 4 shows the coefficients from the first to the fourth principal components. The contribution ratio of the four components was 68.57%. The top 10 players on each principal component score are shown in Table 5.

Looking at the scores of the first component, they are strongly associated with 9 items: “point score”, “kick”, “line break”, “tackle break”, “pass”, “ball touch“, “penalty goals”, and “conversions”. Also the scores of the first component are weakly associated with 3 items: “support”, “tackle assist”, and “penalty”. These 12 items represent the plays relating to the ball. Main positions related the 12 items are SO, CTB and FB, and the players who have high first component score will be SO, CTB and FB. In this analysis, we found that more than a half of the top 10 players in terms of the first component score were foreign players as shown in Table5. From this result, we would say that game controller seems to be chosen from foreign players in JRTL.

The scores of the second component are strongly associated with 5 items: “contact”, “point made”, “off-road”, “tackle breaks”, and “jackal”. These 5 items represent the plays relating to competing possession. Main positions related to the 5 items are FL, No8 and CTB, and the players who have high second component score will be FL, No8 and CTB. Also, more than a half of the top 10 players in the second component were foreign players as shown in Table5. From this result, we would say that players related to competing possession is chosen from foreign players in JRTL.

The scores of third component are strongly associated with 4 items: “penalty goals”, “conversions”, “tackle”, and “assist tackle”. These 4 items represent the plays relating to place kick and tackle. Main positions related to the 4 items are SO and CTB, and the players who have high third component score will be SO and CTB. Now, most of the top 10 players were Japanese players. From this result, we would say that place kicker and tackler tend to be chosen from Japanese players in JRTL.

The scores of fourth component is strongly associated with 2 items: “pass”, and “ball touch”. These 2 items represent the plays relating to passer. Main position related to 2 items is SH, and the players who have high fourth component score will be SH. All of the top 10 players in this component were Japanese players. From this result, we would say that passer is chosen from Japanese players in JRTL.

| Items | First component | Second component | Third component | Fourth component |
|---|---|---|---|---|
| Support | -.802 | .140 | .267 | -.234 |
| Point score | .718 | .145 | .480 | -.396 |
| Kick | .718 | -.090 | .269 | .176 |
| Line break | .596 | .374 | -.310 | -.181 |
| Tackle assist | -.578 | .388 | .512 | .075 |
| penalty | -.550 | .323 | .286 | .155 |
| Turnover lost | .434 | .376 | -.226 | .077 |
| contact | .096 | .888 | -.097 | -.045 |
| Tackle break | .501 | .690 | -.250 | -.081 |
| Point made | -.172 | .643 | .050 | -.060 |
| Offroad | .418 | .557 | -.150 | .228 |
| Jackal | -.331 | .545 | .261 | .099 |
| Turnover won | -.242 | .489 | .221 | .437 |
| Try | .338 | .442 | -.336 | -.091 |
| Penalty goals | .605 | -.037 | .657 | -.389 |
| Conversions | .617 | -.044 | .657 | -.383 |
| Tackle | -.310 | .259 | .615 | .209 |
| Pass | .539 | -.238 | .198 | .725 |
| Ball touch | .603 | -.095 | .252 | .710 |

Table 4. Coefficients for the first to forth principal components referent to each item

| Rank | First component | Second component | Third component | Fourth component |
|---|---|---|---|---|
| 1 | Barnes(SO) | Kikutani(NO8) | Shigemitu(SO) | Hiwasa(SH) |
| 2 | Pisi(SO) | Maekawa(FL) | Tatekawa(CTB) | Tujino(SH) |
| 3 | Goromaru(FB) | Toeava(CTB) | Barnes(SO) | Fukui(SH) |
| 4 | Ogawa(SH) | Latu(NO8) | Monji(SO) | Hamazato(SH) |
| 5 | Shigemitu(SO) | Leitch(FL) | Mitomo(CTB) | Tanaka(SH) |
| 6 | Daniel(FB) | Tatekawa(CTB) | Goromaru(FB) | Nishibashi(SH) |
| 7 | Tamura(CTB) | Arai(FL) | Tamura(CTB) | Namekawa(SH) |
| 8 | Gerrard(FB) | Joe Iongi(NO8) | Jantjies(SO) | Yamamoto(SH) |
| 9 | Viljoen(FB) | Paea(CTB) | Daniel(FB) | Motegi(SH) |
| 10 | Tatekawa(CTB) | Tui(NO8) | Ogawa(SH) | Umeda(SH) |

□Japanese ■Foreigner

Table 5. The top 10 players in each principal component score

The players explained by these four components were divided into six clusters by the cluster analysis. The number of players in each cluster is shown in Table 6, according to their positions. Table 7 also shows the average of the principal component scores of each cluster.

We here explain the characteristics of clusters 1, 2 and 3. In cluster 1, the players who have high score of second component seem to gather as shown in Table 7. That is, mainly FL and NO8 players are in this cluster. In cluster 2, the players who have high score of third component seem to gather. That is, most of FW players are in this cluster. In cluster 3, the players who have high score of third component seem to gather, that is, place kicker is in this cluster.

As seen from the result of principal component analysis the first component seems to represent general tendency in terms of plays relating to the ball, and the second to the forth components are linked to the particular positions which correspond to each cluster. Here, we note that there are some exceptions in each cluster. For example, Colin Bourke is a SO player in cluster 1 shown in Table 6. Here, SO has a connecting role with the FW and BK (Parsons and Hughes, 2001), and cluster 1 consists of mainly FL and NO8, not SO. Colin Bourke has a characteristic of play with competing possession rather than pass. That is, his play style as a SO seems to be different from other SOs. As Colin Bourke is foreign player, he might have a role of competing possession in his team. We can infer this kind of insight just from the frequency data.

As seen from the result of principal component analysis, the third component linked to FW which correspond to cluster 2. There are also some exceptions. For example, the CTB player (Masatoshi Miyazawa) was in cluster 2. He has a characteristic of play with tackle than ball carry, while the other CTB players make different play styles. He is not larger than other CTB players (height 5 feet 6.93 inches and weight 176.37 lbs.). Possibly, he is a key player of defense in his team. In addition, Masatoshi Miyazawa is Japanese player and may play a role of tackler.

According to James et al. (2005), any positions played by any players who have different playing styles. However, the roles of Japanese and foreign players may be decided in JRTL, that is, Japanese play as tackler and passer, and as foreign players as game controller and competing possession player.

| | cluster 1 | cluster 2 | cluster 3 | cluster 4 | cluster 5 | cluster 6 |
|---|---|---|---|---|---|---|
| PR | 0 | 19 | 0 | 0 | 0 | 2 |
| HO | 1 | 5 | 0 | 0 | 0 | 4 |
| LO | 4 | 14 | 0 | 0 | 0 | 5 |
| FL | 10 | 11 | 0 | 0 | 0 | 1 |
| NO8 | 7 | 4 | 0 | 0 | 0 | 3 |
| SH | 0 | 0 | 1 | 0 | 11 | 0 |
| SO | 1 | 0 | 5 | 0 | 4 | 0 |
| CTB | 2 | 1 | 5 | 5 | 0 | 10 |
| WTB | 0 | 0 | 0 | 7 | 0 | 14 |
| FB | 0 | 0 | 3 | 5 | 0 | 7 |
| Total | 25 | 54 | 14 | 17 | 15 | 46 |

Table 6.Number of players in each position in each cluster

| component/cluster | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| First component | -0.35 | -0.95 | 1.96 | 1.09 | 0.88 | 0.06 |
| Second component | 1.40 | -0.31 | -0.12 | 0.89 | -1.00 | -0.47 |
| Third component | 0.13 | 0.35 | 2.21 | -1.31 | 0.18 | -0.75 |
| Fourth component | 0.38 | -0.15 | -1.16 | -0.24 | 2.46 | -0.40 |

Table7. The average of the principal component scores of each cluster

## 4. CONCLUSION

In this study, we tried to evaluate the teams and the players in JRTL, comparing them with world level teams and players. As the result, in terms of team evaluation, the defense of the teams in JRTL seems to be weaker than that in SNT. In terms of player evaluation, most of the passers and place kicker as tackler are Japanese players in JRTL, although few Japanese players are game controller and competing possession players. Japanese players seem to be weak in tackle, game control and competing possession. Therefore, JRTL would need to foster Japanese tackler, game controller and competing possession player to strengthen the Japan national rugby team.

## References

James, N., Stephen, D. M, & Jones, M, P, N. (2005). The development of position-specific performance indicators in professional rugby union. Journal of Sport Sciences, 23, 63-72.

Parsons, A., & Hughes, M, D. (2001) Center for Performance Analysis, U.W.I.C., Cardiff, CF23,6,129-136

# ON THE DISTRIBUTION OF WINNING MARGINS IN RUGBY LEAGUE

Elliot Tonkes [a]

[a] *Queensland University of Technology*

*Corresponding author:* Elliot.Tonkes@qut.edu.au

## Abstract

The winning margin in Rugby League football games has importance from several perspectives, including the opportunity to win 'bonus' points in certain league competitions and in betting forums which accommodate margin wagering. The statistical distribution of final scores and winning margins exhibits a detailed structure with discreteness inherited from the different scoring mechanisms. Historical data from bookmakers and match outcomes in 2009-2015 NRL are analysed to confirm that both historical scoring outcomes and gambling odds reflect the discreteness of scoring channels. Using said data, we derive and calibrate a probability distribution for scores and margins constructed from a model using Poisson processes to describe scoring events. The model predictions are compared with real outcomes to assess whether scoring processes for each team can be treated as independent processes, or whether the nature of the game induces dependence into the respective teams' scoring. The model is endowed with time- and score-dependent intensity rates to incorporate the effect of gameplay strategies.

**Keywords**: Rugby League, Poisson Process, Scores, Betting Odds

## 1. INTRODUCTION

The Australian National Rugby League (NRL) draws 2 million viewers weekly and constitutes a significant proportion of Australia's annual sports betting expenditure of $3 billion [Wallace, 2016]. Each game of Rugby League ('League') is played for two 40 minute halves and sees numerous gambling agencies offering around 200 distinct betting products, with the payout on many of the products dependent on the full time score. The dominant wagers are head-to-head (H2H) bets which are decided by the team with the most points at full time.

League has particular scoring channels which yield interesting patterns for final scores. The scoring system [NRL 2014] has: Unconverted Try (UT) (4 points) Converted Try (CT) (6 points), Penalty Goal (PG) (2 points) and Field Goal (FG) (1 point). Tries are the most common form of scoring, penalty goals are less common and field goals relatively rare.

The objective of the current work is to establish a parsimonious model to describe the probability mass function (*pmf*) of full-time winning margins in a given game of League. There are three particularly interesting features of the *pmf* that are observed in both historical statistical distributions and bookmakers odds, namely:

(i) the expected winning margin holds a relationship with head-to-head betting odds (Figure 1): in other words a team which is at short odds to win will statistically win by a large margin;

(ii) there appears an interesting 'micro-structure' in the *pmf* governed by the discreteness of scoring events (Figure 5): the appearance in League of a winning margin of 3 points is less likely than 4;

(iii) there is a negative correlation exhibited between home and away scores in League games, which influences the *pmf* (Figure 5).

Seminal work [Lee, 1999] developed a model with negative binomial distributions and a copula technique to address the issues of significant covariance between home and away scores. The study focussed on CT, UT and PG scoring with FG being treated with a homogeneous Poisson process. The models contained numerous team-specific parameters to quantify offensive and defensive properties of teams and compared outcomes of the 1996 season.

Rugby league betting markets in England have been assessed [Simmons et al, 2014] and the relationships with home advantages assessed. The information in betting markets has been used in forecasting the outcomes of football matches, for H2H competitions [Dobson, 2007] as well as likely spreads [Gandar, 1998].

Clarke [Clarke 1992] analysed the distribution AFL winning margins as a function of tipping preferences. Glasson [Glasson 2006] used Brownian motion to model the stochastic movements of AFL scores. Poisson and negative binomial distributions have been used to model soccer scoring [Croucher 2000 and references therein].

In the current work, we seek a model conditioned only on the information contained in the H2H odds. We assert that subtle game influences such as home-advantage, key player influence, offensive and defensive strengths of individual teams are contained in the H2H odds. Our innovation for League is to introduce a nonhomogeneous process for the FG scoring channel consistent with observed strategic behaviour. Intuitively when a game sits with certain critical margins, there exists an incentive to attempt FG. We also postulate that there is no need for the model to introduce additional bivariate structure into the distribution of home and away scores, because the H2H odds already contain such information.

## 2. MODEL PRESENTATION

For a particular game, at any time $t$ up to full time $T$, denote the score of the home team as $H(t)$, the away team as $A(t)$ and the margin $M(t) := H(t) - A(t)$. The scores are generated by the numbers $n$ of CT, UT, PG and FG events with score $6\, n_{CT} + 4\, n_{UT} + 2\, n_{PG} + 1\, n_{FG}$ for each of the home and away teams.

In the standard way, define the intensity of a Poisson process as $\lambda(t)$, meaning that in time $(t, t+dt)$ the probability of an event is $\lambda(t)dt$. For a given game denote the probability for the home team to win at time $T$ as $W$ inferred from (normalised) bookmakers' H2H odds.

Define the following margin scores as 'special' levels: $H^* = \{-1, 0, 6, 12, 18\}$, $A^* = \{-18, -12, -6, 0, 1\}$, and denote the following 'special' periods: $\tau_{half} = (35,40]$, $\tau_{full} = (60,80]$ when the intensity of FG is elevated.

Define the process for CT, UT, PG and FG as independent Poisson processes with intensities described with a game-specific factor $\alpha$ derived in (3). Here CT, UT and PG are homogeneous processes. FG is nonhomogeneous with dependence on time $t$ and the prevailing margin $M(t)$.

$$\begin{aligned}
\lambda^H_{CT} &= \alpha\, \lambda^0_{CT} & \lambda^A_{CT} &= (2-\alpha)\, \lambda^0_{CT} \\
\lambda^H_{UT} &= \alpha\, \lambda^0_{UT} & \lambda^A_{UT} &= (2-\alpha)\, \lambda^0_{UT} \\
\lambda^H_{PG} &= \alpha\, \lambda^0_{PG} & \lambda^A_{PG} &= (2-\alpha)\, \lambda^0_{PG} \\
\lambda^H_{FG}(t,M) &= \alpha\, \lambda^{H0}_{FG}(t,M) & \lambda^A_{FG}(t,M) &= (2-\alpha)\, \lambda^{A0}_{FG}(t,M)
\end{aligned} \tag{1}$$

The specific function for the intensity of process FG is:

$$\lambda^{H0}_{FG}(t,M) = \begin{cases} \lambda^0_{FG} \text{ if } t \in \tau_{half} \text{ AND } M \in H^* \\ \max[0, 0.1 \times \lambda^0_{FG}(t-60)] \\ \qquad \text{if } t \in \tau_{full} \text{ AND } M \in H^* \\ 0 \text{ otherwise} \end{cases} \qquad \lambda^{A0}_{FG}(t,M) = \begin{cases} \lambda^0_{FG} \text{ if } t \in \tau_{half} \text{ AND } M \in A^* \\ \max[0, 0.1 \times \lambda^0_{FG}(t-60)] \\ \qquad \text{if } t \in \tau_{full} \text{ AND } M \in A^* \\ 0 \text{ otherwise} \end{cases} \tag{2}$$

The scalar $\alpha$ is established by forcing consistency of model and bookmakers' probabilities:

$$\Pr(M(T)>0) = W \tag{3}$$

where $M(T)$ is the full-time margin score achieved by the process governed by (1) and (2).

## 3. JUSTIFICATION FOR THE MODEL

We analyse data from historical games to confirm suitability of the assumptions underlying the Model (1-3).

*1. Head-to-head odds as unbiased predictors.*

Model (1-3) incorporates H2H odds to inform the intensity for each of the home and away scoring processes. It is well understood that (synthetic) probabilities implied from betting markets do not necessarily reflect the real-world probabilities for game outcomes [Kuypers, 2000]. However, a quantitative analysis of historical odds for NRL gambling (Figure 1) using data from [ASB] supports the hypothesis that the published H2H odds are unbiased. Over $N$ games, the expected return of a bet of $\$1/N$ bet for the home team to win each game is \$1 under assumption that the odds are fair (and bookmaker profit is neglected). By conducting many Bernoulli experiments using the stated odds (adjusted to remove bookmaker profit per section 4), we are able to establish a confidence interval (CI) for the returns.

| Season | Return on \$1 (90% CI) |
|---|---|
| 2009 | 1.11 (0.87,1.10) |
| 2010 | 1.02 (0.89,1.12) |
| 2011 | 1.08 (0.89,1.12) |
| 2012 | 1.06 (0.88,1.12) |
| 2013 | 1.06 (0.88,1.12) |
| 2014 | 1.07 (0.88,1.13) |
| 2015 | 0.95 (0.88,1.12) |
| ALL | 1.04 (0.95,1.05) |

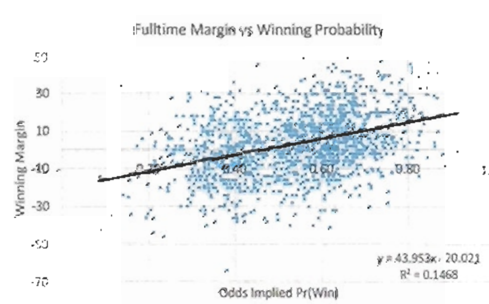



Figure 1: Left: Fairness of H2H home-team odds in NRL 2009-2015, Right: Winning margin dependence on bookmaker H2H odds (as implied probability of winning) 2009-15

One intuitively expects that higher probabilities of winning will on average result in progressively higher winning margins. Figure 1 illustrates for all NRL games over 2009-2015 a comparison between the implied probability of winning and the actual winning margin (negative for losing). One may immediately discern that the relationship is heavily endowed with noise but ANOVA confirms that betting odds provide significant explanatory power ($p$ value $10^{-11}$).

2. *Poisson process for scoring: homogeneity and independence*

Fundamental characteristics of Poisson processes include the independent nature of events and the commensurate *pmf* of the numbers of events per game [Grimmett, 1992] via equation (4). Analysis of minute-by-minute scoring data from all 201 games in the 2015 NRL season established the times of scoring events, bucketed in 5-minute intervals (Figure 2). It shows that FG events occur under the dual conditions of (i) time in 'special' periods; and (ii) the margin in 'special' levels. We find that the subtle requirement to model FG with this dependency is critical for an accurate *pmf*. On the other hand UT, CT and PG remain homogeneous.

The FG intensity is motivated by circumstances familiar to League observers. The FG event is known to become extremely important as a mechanism to either (i) provide a tie breaker or (ii) consolidate a winning score by placing a team into a position which is at least $N$ scoring events ahead of the opposition or (iii) recover a position from one-point-down to recover a draw. It may be quickly enacted if there is insufficient time before half/full time to set up a try.

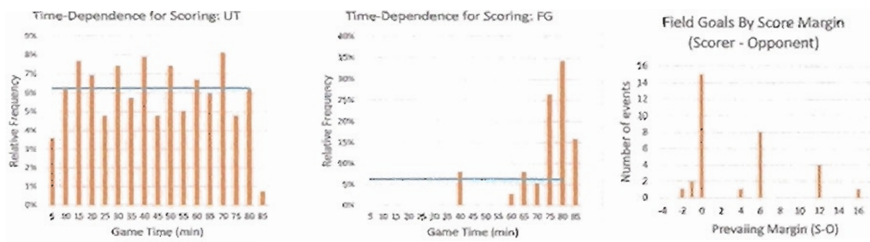



Figure 2: Left: homogeneity of UT Events. Centre: Dependence of FG on Time, Right: Dependence of FG on prevailing Margin

We make reference to the form of the factors $\alpha$ and $(2-\alpha)$ in expression (1). It is well known that the sum of two Poisson processes with intensities $\alpha\lambda$ and $(2-\alpha)\lambda$ yields a Poisson process with intensity $2\lambda$, with times between events governed by an exponential distribution [Grimmet, 1992]. Consequently, under an assumption that all games are governed by a common scoring intensity for CT, UT and PG, the parameters $\lambda^0_{CT}, \lambda^0_{UT}, \lambda^0_{PG}$ can be established from the total scoring intensity across all games. Figure 3 shows the distribution of numbers of tries per game and also the times between tries and compares each to Poisson and exponential distributions with the values of $\lambda^0_{CT}$ and $\lambda^0_{UT}$ calibrated in section 3.

$$\Pr(N=n) = (\lambda T)^n \exp(-\lambda T)/n! \qquad (4)$$

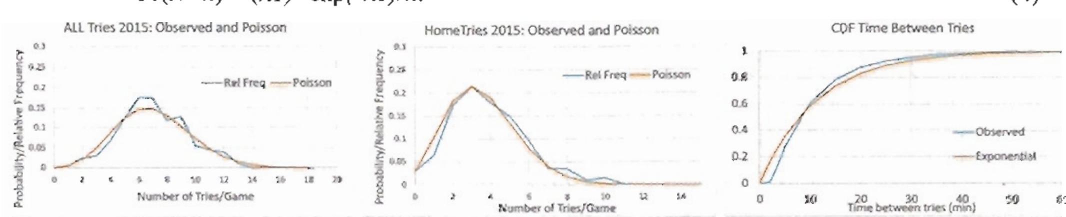



Figure 3: Left/Centre: Relative frequency of all/home observed tries per game over 2015 season and Poisson *pmf*, Right: Times between tries and exponential density calibrated per section 3,

3. *Explanation of negative correlation and overdispersion*

It is well documented [Lee, 1999] that Rugby League Home and Away scores exhibit a negative correlation and also possess overdispersion compared to a Poisson distribution fit. We provide an explanation regarding how Model (1-3) addresses each of these characteristics.

The scattergram in Figure 6**Error! Reference source not found.** exhibits instances of high home/low away scores and vice versa, but rarely a high score for both teams. Model (1-3) presumes that all games exhibit a common scoring intensity, and that some of the intensity is transferred from the weaker team to the stronger to a degree predicted by bookmaker odds, but in a way that preserves the total expected number of scoring outcomes in each game. This has the effect of biasing outcomes to the top left or bottom right of the scattergram and induces correlation.

Under Model (1-3) each (say) home score is not drawn from the same distribution. Instead, the home scoring processes are endowed with a different mean for each game which is influenced by the H2H odds, and leads to a wider spread of scoring outcomes, thus generating the overdispersion compared to drawing all outcomes from a common Poisson distribution.

## 3. CALIBRATION OF THE MODEL

The Model (1-3) has been calibrated to observed data in the 2015 season of the NRL [ASB, MatchCentre], with estimation of intensity parameters $\lambda$ generated by (total events) / (total time). The following parameters arise from the regular time outcomes over 201 games (16,080 minutes).

| Quantity | Observations | Derived Quantity | Value |
|---|---|---|---|
| Number of CT | 1004 | $\lambda^0_{CT}$ | 0.0312 per min |
| Number of UT | 415 | $\lambda^0_{UT}$ | 0.0129 per min |
| Number of PG | 201 | $\lambda^0_{PG}$ | 0.0063 per min |
| Number of FG | 32 | N/A | 0.001 per min |

Table 1: Model parameters derived from 2015 season.

We calibrate the FG model (2) by deriving the conditional intensities. Game times are divided into intervals $\tau_1 = (0,5]$, $\tau_2 = (5,10]$, … $\tau_{15} = (75,80]$, and margins $M = -60, .., 60$. Across all games in a season, we calculate $\mathscr{T}(\tau,M)$ as the total number of minutes spent in each interval $\tau$ and at each margin $M$. We then calculate the total number of field goals $\mathscr{N}(\tau,M)$ executed in each state, and derive the intensity $\Lambda(\tau,M) = \mathscr{N}(\tau,M) / \mathscr{T}(\tau,M)$. The observations are fitted as illustrated in Figure 4 and we arrive at $\lambda^0_{FG} = 0.0216$ per min.

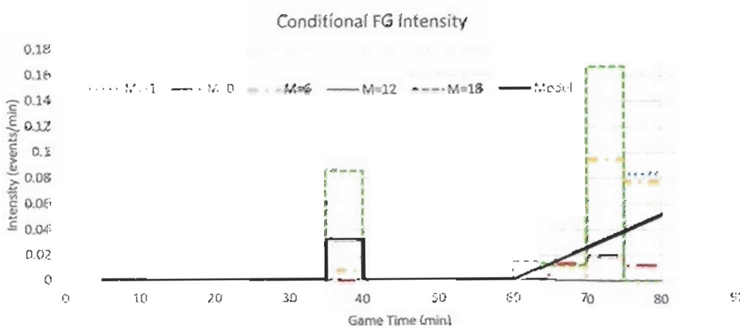



Figure 4: Observed intensity for field goals per minute conditioned on $t$, $M$; and fitted model

## 4. MODEL IMPLEMENTATION

The model has been implemented in Matlab by stepping forward over 5-minute time increments and updating the score *pmf*. The initial distribution is known ($M = 0$ with probability 1). At step $j$, the density is known for time $t = 5(j-1)$. Over a 5-minute increment, the general formula (4) is applied to derive the probability of home and away occurrences of CT, UT, FG and PG, with intensities governed by expressions (1-3). The result is used to update the distribution for time $t = 5j$. The algorithm generates results almost instantaneously and Monte Carlo simulation is not required.

The method to solve (3) for $\alpha$ has been implemented by a numerical technique. The *pmf* for $M(T)$ has been derived from (1), (2) using a fine grid of $\alpha$ values. For each $\alpha$, we can calculate the probability $\Pr(M(T)>0)$ directly, and this provides a map between $W$ and $\alpha$ via piecewise linear interpolation.

Observed bookmaker odds contain a profit. To infer the implied probabilities, we have scaled to normalise odds [Kuypers, 2000]. That is, if mutually exclusive events $e_1, .., e_n$ span all possible outcomes with payoffs $q_1, .., q_n$ then the implied probability of $e_1$ is $1/q_1 \times 1/(1/q_1+1/q_2+..+1/q_n)$. This technique is applied to the H2H odds to derive $W$, and to establish the implied probabilities of margins derived from exact-margin markets.

## 5. RESULTS AND DISCUSSION

### COMPARISONS OF HISTORICAL AND BOOKMAKER DISTRIBUTIONS

We generate a *pmf* for each historical game over 2009-2015, with $\alpha$ calibrated to the H2H odds quoted for contest. An aggregate *pmf* is then established for the entire season by equally weighting each game. Of course, this does not represent a distribution for a given game, but provides a mechanism to compare accuracy of the overall model. The distribution derived by this method is compared to the distribution of actual game outcomes, and standard presentation methods are used to make the comparison, namely comparison of the *pmf*, CDF and *p-p* plots. Because the even and odd scores have such disparate structures, we also present the CDF for even and odd scores separately. For brevity, we present the results for 2009-2015 in aggregate, but in-sample and out-of-sample results for individual seasons show similar characteristics.

We compare the model *pmf* against probabilities implied from the Sportsbet's exact-margin markets (allowing bets on any margin between -30 and +30 points) for two typical matches in the 2016 season (Panthers paying \$2.20 and Sharks paying \$1.30).

The performance of Model (1-3) was compared with variation in which a ***homogeneous*** FG process is used where an intensity of 0.001 FG/min (Table 1) is applied. The homogenous model results exhibit a smoothly varying distribution, rather than the peculiar patterns in Figure 5.

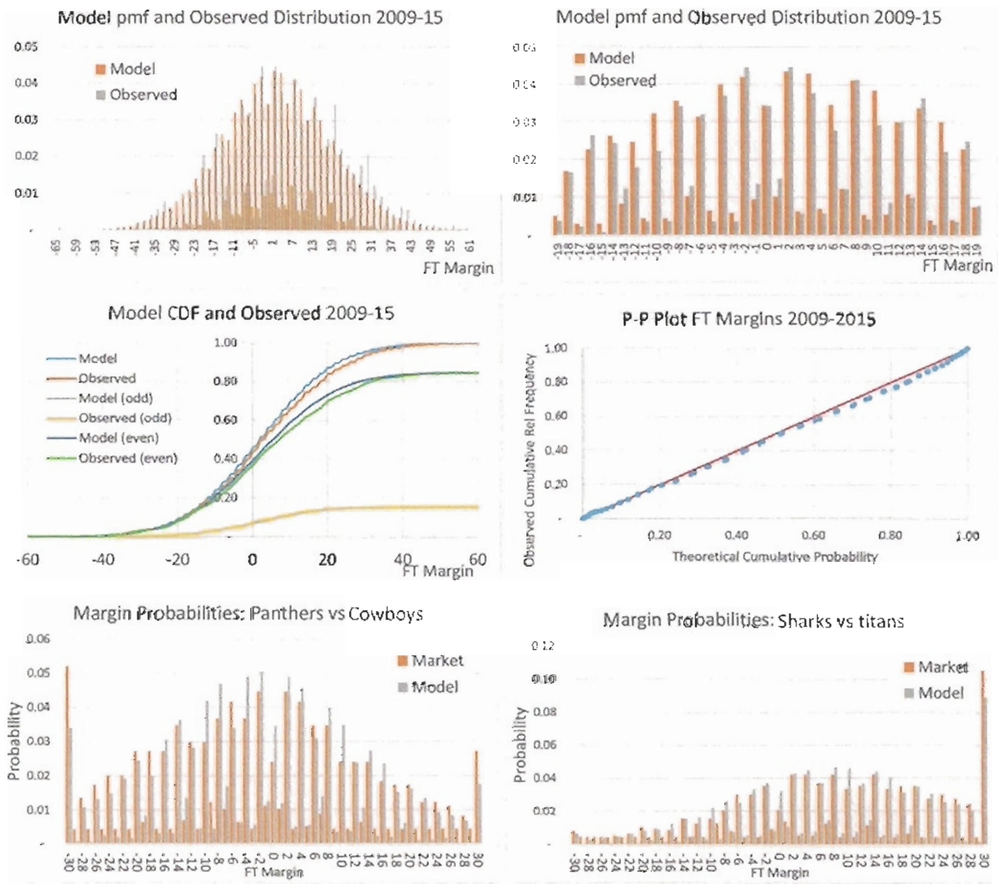
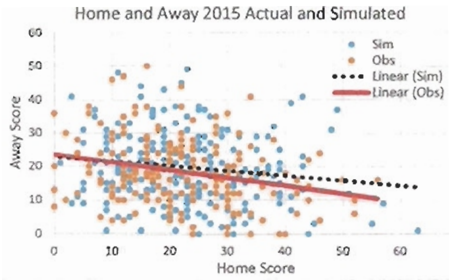



Figure 5: Performance of Model (1-3) to observed outcomes 2009-2015. Top row: *pmf*, Centre row: CDF (all, odd, even) and *p-p* diagram of historical data and *pmf*, Bottom row: *pmf* comparison against Sportsbet odds.

ABILITY TO CAPTURE CORRELATION

Section 2.3 proposes that the overdispersion and negative correlations are driven by the $W$ dependence of distributions which govern each individual game. We are able to assess the accuracy of that assertion by establishing a theoretical correlation which is induced by the H2H odds and compare to actual outcomes from season-to-season. For illustration, we have generated a simulation of home and away outcomes arising from the 201 games from season 2015 consistent with the method in Section 4 and compared the covariance structure against the actual results in Figure 6.

It is found Model (1-3) yields a correlation of around −0.15, which is a weaker relationship than the actual correlation of −0.26. However, the evidence from our principal results in Figure 5 indicate that this fact is not deleterious to the model performance. Statistics are presented in Figure 6 for all games over 2009-15 but similar outcomes arise for individual seasons, (with the exception of better accuracy for the 2015 in-sample period).




| Statistic | Observed | Model |
|---|---|---|
| Correlation (H/A) | -0.30 | -0.15 |
| Mean (H) | 22.35 | 21.31 |
| Std (H) | 11.29 | 11.11 |
| Mean (A) | 18.57 | 19.07 |
| Std (A) | 10.24 | 10.55 |
| Mean (M) | 3.77 | 2.24 |
| Std (M) | 17.41 | 16.48 |

Figure 6: Comparison of actual game negative correlation and Model (1-3). Actual and model statistics on Home, Away and Margin scores.

## 6. CONCLUSIONS AND FUTURE WORK

The Model (1-3) is a simple model based on independent Poisson processes and performs admirably to generate distributions for winning margins consistent with historical data and observed bookmakers odds. The inclusion of H2H odds to bias the intensities of teams' scoring efforts is suffieient to induee dispersion and to some extent negative correlation in home/away scoring. The model contains nonhomogeneous processes, but the Markov property of the model means that the implementation is straightforward. The model is easy to calibrate and the derived *pmf* exhibits key statistical characteristics, including different attributes for odd and even scores. We identify that home-team bias is already captured in the model via the H2H odds.

Model (1-3) does not accommodate a feature observed in the real world, whereby a 1-2 minute delay occurs between scoring a try and the conversion attempt, and the discrepancy is apparent in Figure 4 near the origin. An improvement to the model could incorporate this lag.

There are implications arising from this study for scoring systems and extra time. The particular FG process is governed by special scores and times which are manifested by teams acting in a strategic manner to win. If the scoring outcomes took on different values (for example rule changes to make field goals worth more, or Rugby Union scoring systems), then the strategies would adapt and the processes described here would not apply. It would be interesting to establish if the intensities derived here from revealed behaviour actually represent the optimal times and scores to focus on field goals.

### References

ASB, Australia Sports Betting Data Repository: http://www.aussportsbetting.com/data/historical-nrl-results-and-odds-data/ accessed 10 March 2016

Croucher, J., (2000), "Analysing Scores in English Leagne Soccer", in Cohen et al (eds). Proceedings of the 6th Mathematics and Computers in Sport", Sydney, pp 119-126

Clarke, S. (1992), Compnter and human tipping of AFL football – a comparison of 1991 results", in de Mestre (ed), Proceedings of Mathematics and Compnters in Sport, Bond University, pp. 81-93

Dobson, S., Goddard, J. (2007). Forecasting scores and results and testing the efficiency of the fixed-odds betting market in Scottish league football, in Albert,J., Koning, R.H. (eds), Statistical Thinking in Sports. Chapman and Hall, pp. 91-109

Gandar, J., Zuber, R., O'Brien, T., Russo, B., (1998), Testing Rationality in the Point Spread Betting Market, Journal of Finance, *43(4)*, pp. 995-1008.

Glasson, S., (2006), A Brownian motion model for the progress of Australian Rnles football scores, in Hammond et al (eds), Proceedings of the 8th Conference on Mathematics in Sport, Bond University, pp216-225.

Grimmett, G.R. and Stirzaker, D.R., (1992), Probability and Random Processes 2nd ed, Oxford.

Joan Hu, X. and Lawless, J.F. (1996), Estimation of Rate and Mean Functions from Truncated Recurrent Event Data, J. American Statistical Association, 91(433), pp.300-310

Kuypers, T., (2000) Information and efficiency: an empirical study of a fixed odds betting market, Applied Economics, *32(11)*, pp. 1353-1363

Lee, A. (1999). Modelling rugby league data via bivariate negative binomial regression. Aust N Z J Stat. *41*, pp. 141–152.

MatchCentre, NRL Match Centre Data Repository, https://matchcentre.nrl.com/, accessed 10 April 2016.

NRL, (2014) 2014 NRL LAWS & INTERPRETATIONS, https://www.nrl.com/portals/nrl/RadEditor/Documents/2014NRLLawsandInterpretation.pdf accessed 10/5/2016

Simmons, R., Forrest, D. and Curran, A. (2014), "Efficiency in the Handicap and Index Betting Markets for English Rugby League", in "The Economics of Gambling", Vaughan-Williams, L. Routledge, pp115-134

Wallace, R, (2016), "Everything to play for in sports betting", The Australian, 27 Jan 2016

# Forecasting in Rugby Union: A Comparison of Methods

William Undy [a,b,d] , Dion R.J. O'Neale [a,b] , Ken Quarrie [c]

[a] *The University of Auckland*
[b] *Te Pūnaha Matatini – a New Zealand Centre of Research Excellence*
[c] *New Zealand Rugby*
[d] *Corresponding Author: wund525@aucklanduni.ac.nz*

## Abstract

Many people want to predict how a sports team will perform in a given season. Fans, analysts, and sports bettors all have an interest in pre-season predictions and their accuracy. There have been many methods developed to predict team performance in many sporting codes. However, rugby union has had relatively little research into the efficacy of different methods to predict team performance. In this research we applied five different predictions methods to Super Rugby. These methods encompassed some of the major ideas used in prediction; simplistic result-based rating systems, points scored and conceded, and team chemistry – quantified using a time-varying multi-partite network. First, we adapted these methods so they could be used for Super Rugby predictions. We then evaluated the methods by comparing them on a variety of different metrics. The results of the research are interesting because they not only provide a talking point for Rugby prediction but form a basis from which more complex methods can be developed.

*Keywords*: Rugby union, forecasting, multi-partite networks

# SIMPLIFYING PROBABILISTIC MATCH IMPORTANCE CALCULATION IN FOOTBALL

Michael de Lorenzo [a, b], Bradley O'Bree [a], Ian Grundy [a]

[a] *School of Science, RMIT University, Melbourne, AUS*
[b] *Corresponding author: s3285802@student.rmit.edu.au*

## Abstract

Assessing match importance in professional football (soccer) is beneficial for a number of reasons, primarily in predictive modelling for match outcome and attendance. Previously, modelling of match importance has focused on some specific end-of-season aim, such as winning the league championship or avoiding relegation. Calculation has typically involved the use of retrospective measures or complex computer simulation procedures. These methods have drawbacks, where retrospective measures cannot be implemented into a live season; and complex computer simulation procedures require a great deal of runtime. In this paper, we explore the effect of simplifying the calculation of match importance using a probabilistic method by relaxing match outcome independence assumptions at varying degrees of severity in an attempt to counteract these drawbacks. The probabilistic measure builds off of previous research by defining match importance as the difference between success probabilities conditional on the result of a match. Probabilities follow a trinomial distribution to account for the possibility of each team winning, losing or drawing a match; and for simplicity remain constant throughout a season and independent of the teams playing. A complete season simulation model provided the basis for comparison with the simplified model, where analysis compared the relative size of the bias of outcome probabilities and the characteristics of importance distributions for teams aiming to finish in the top position at the end of the season. Results indicated that the complexity of the complete simulation procedure is not required as the distribution of importance remains similar to that of the simpler probabilistic measure. Results also suggested that team strength should be incorporated into match importance calculations because the assumption of constant match outcome probabilities between teams underestimates the variation in round by round cumulative season points totals.

**Keywords:** Match importance, trinomial distribution, simulation, football

## 1. INTRODUCTION

In past research, predictive modelling of match outcome and attendance has included variables such as weather conditions and population of the home team's town (Baimbridge, Cameron, & Dawson, 1996). One common inclusion is the importance of the match for both competing teams, which is often modelled with respect to some specific end-of-season outcome, such as winning the league championship or avoiding relegation. Calculation of match importance has typically involved the use of retrospective measures (Jennett, 1984) or complex computer simulation procedures (Lahvička, 2015). However these methods have drawbacks, where retrospective measures cannot be implemented into a live season, and complex computer simulation procedures require a great deal of runtime.

One of the most popular retrospective match importance measures for football was conceived by Jennett (1984). The author defined match importance by first establishing the points total of the league champion at the end of the season, and then applying the result retrospectively throughout the season to determine the required wins for a team to achieve the points total. This method has been adapted into Australian Rules football (Borland & Lye, 1992) and English one-day cricket (Morley & Thomas, 2005). A similar approach to retrospective measures is to use a dummy variable (Baimbridge et al., 1996), where a value of 0 indicates a game is not important and a value of 1 indicates a match is important. However, the scale of the match importance is lost in this approach as importance can only take two values. The most common computer simulation procedure applied in past research is the Monte Carlo approach (Lahvička, 2015). The method simulates a season a number of times and determines from the results which matches are important to a team. While this method provides a complete look at a season when determining importance, the key drawback is that it requires a great deal of runtime to complete calculations.

To counteract these drawbacks and simplify the match importance calculations we propose a probabilistic method. Probabilistic measures of match importance have been explored in past research. Schilling (1994), building off of work completed on the importance of points in tennis (Morris, 1977), defined match importance as the difference between two conditional probabilities: the probability that a team wins a best-of-seven game series given they win the next game, minus the probability that the team wins the series given they lose the next game. This definition has been applied to Australian Rules football (Bedford & Schembri, 2006), NBA basketball (de Lorenzo & Bedford, 2014) and combined with Monte Carlo simulation in English Premier League football (Scarf & Shi, 2008). The probabilistic method we propose builds off of the previous research by introducing a trinomial distribution function to account for drawn matches. We compare the method to a pure Monte Carlo simulation model to determine if the importance distributions change despite the reduction in calculation complexity.

## 2. METHODS

The definition of match importance for football follows the work completed by Schilling (1994) by using the difference between conditional probabilities. However, instead of assessing importance with respect to winning a seven game series, we calculate importance with respect to winning the league championship (finish first place) in German Bundesliga football. While there are other end-of-season outcomes in Bundesliga football (European cup qualification, relegation), we elect to focus on winning the league championship with the intention to expand to include other outcomes in future research.

The conditional probabilities definition by Schilling (1994) is calculated by assessing the difference between the probabilities of winning the series given the team wins or loses the next game. However, in football there are three match outcomes (win/draw/loss) instead of two (win/loss), where a draw outcome can still be beneficial for a team despite being less desired than a win. Therefore, we calculate win and draw importance using the conditional probabilities and combine the two to generate a result importance. Result importance is defined as the difference between two probabilities: the probability that a team wins the league championship given they achieve a non-negative (win/draw) result in their next match (g+1); minus the probability that a team wins the league championship given they achieve a negative (loss) result in their next match.

(1)

A Monte Carlo simulation model provides the basis for comparison with the simplified model. Developed using a custom VBA program in Microsoft Excel, the procedure first generates a season schedule following the format used in Bundesliga football, where each team plays each other twice over 34 rounds. Each generated season schedule is then simulated to completion a number of times to generate success probabilities (finish first) dependent on individual match outcomes throughout the season. Match importance is then evaluated using these probabilities conditional on the current round and the team’s current position.

To determine the outcome probabilities for an individual contest, match results from seasons 2005/2006 through 2014/2015 for first and second division German Bundesliga were collected from www.football-data.co.uk., totalling 5,508 matches. Over the span of the nine seasons, 26% of matches were found to be drawn, meaning 74% of the time a victor was decided. This 74% is split evenly between win and loss to create win-draw-loss probabilities of 37%-26%-37%, respectively. A drawback of using these constant probabilities for all teams is that team strength is ignored; where a strong team should have a larger win probability given they are a dominant team. Despite this drawback, we elect to use the constant probabilities to create a baseline model, which can be expanded in future research through variation of the probabilities.

The simplified probabilistic model calculates the match importance using a trinomial distribution. This is an extension of the method applied by Bedford and Schembri (2006) but replaces a binomial distribution with the trinomial in order to take into account the three match outcome probabilities. Unlike the Monte Carlo simulation model, the simplified probabilistic model only requires the points of the league leader and the points of the current team of interest. By identifying success and failure scenarios such that the lower ranked team can overtake the higher ranked team, the trinomial distribution calculates the probability that a team can win the league championship given they win, draw, or lose their next match. In order for the trinomial distribution to calculate the conditional probabilities and therefore the importance, match outcome probabilities are required. For this, the constant probabilities outlined in the Monte Carlo simulation model are used. Like the Monte Carlo model, the simplified trinomial model calculates both a win importance and draw importance before combining to the two to generate a final result importance.

The trinomial distribution simplified model is compared with the simulation model using both simulated and observed season results. For the Monte Carlo model, 10,000 iterations are applied, with a further 100 season result iterations to generate the outcome probabilities. While the 100 iterations may seem low, it was found through trial and error that a larger number of iterations dilate the outcome probabilities. For the trinomial simulated seasons, the schedule is generated and results are produced similar to the Monte Carlo method. Once the results have been created, the trinomial distribution is applied and the importance of each match throughout the season is calculated. For the observed seasons, the trinomial distribution is applied to the results from actual seasons of division one and division two Bundesliga football. For the two divisions, the win/draw/loss frequencies from each specific division are used when calculating the conditional probabilities.

## 3. RESULTS

The analysis commences with an exploration of the average match importance across all teams after each round of the season, with the aim being to determine if the distribution of average importance throughout a season differs between models. Figure 1 displays the average match importance throughout a standard 34-round season, and includes observed results from both division one and division two Bundesliga.

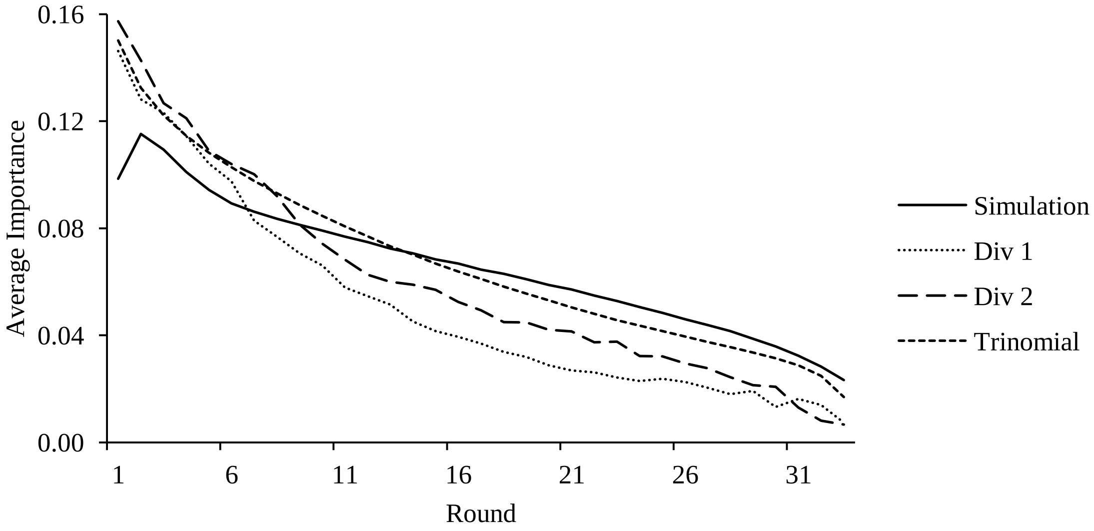



Figure 1: Average match importance for simulation and observed seasons

As observed in Figure 1, the average match importance across all four models follow a similar path, with average importance peeking at the start of the season before steadily decreasing as the season progresses. This is an expected result as there would be less teams competing for the league championship at the end of the season, meaning the average importance across all teams would be low. Despite some noise at the start of the season, the two simulation models share a similar shape and include a small drop at the conclusion of the season. The closeness of these two distributions suggests that the reduction in computational complexity implemented in the trinomial model has produced a similar result to the Monte Carlo simulation approach.

While the average importance distributions have shown that all four models are similar, an exploration of importance by specific positions helps to provide more information about the reduction in computational complexity. The importance distributions for positions 1, 6, 13 and 18 for both the Monte Carlo simulation and the division one observed data are presented in Figure 2. The observed trinomial model is presented here in place of the trinomial simulation as the former is seen to follow a similar distribution as the latter in Figure 1, albeit slightly lower. We discuss this small difference later in this section.

As seen in Figure 2, there are some differences and similarities with the distribution of importance by specific position. For position 1, the observed curve shows a steady increase throughout the season before declining while the simulation model curve shows a dramatic increase at the start before a slow steady decline around the same time as the observed curve. For position 6, both models show a steady decline as the season progresses, despite the fact that they are mirroring each other. Finally, positions 13 and 18 provide almost identical importance distributions, with the importance being high at the start of the season before declining. Overall, the results for both models show what would be expected, with the lower positions showing importance decreasing rapidly early in the season while the higher positions, who are in the running to claim first place, show importance grows as the season progresses. Despite some positions differing slightly, the results are promising and support the theory that the trinomial approach provides similar results to the Monte Carlo, without requiring a long run time or complex computer calculations.

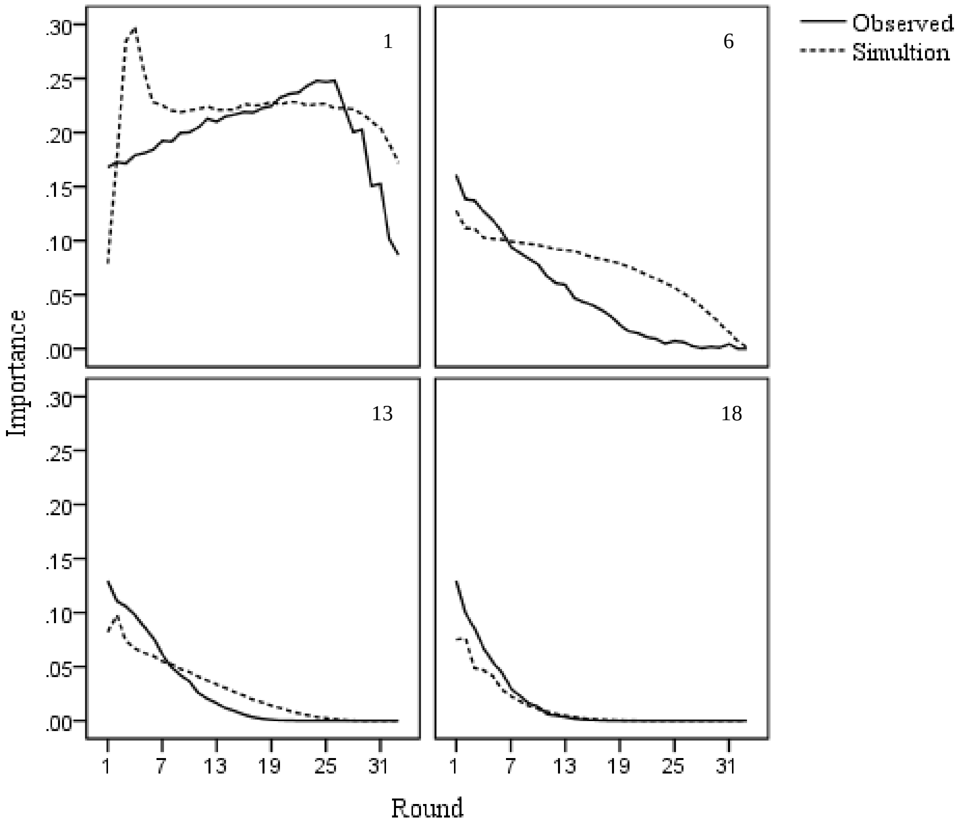



Figure 2: Position Specific Importance for Simulation vs. Observed (Division 1)

As previously mentioned, one result observed in Figure 1 was that the observed trinomial distributions were slightly lower than the trinomial simulation model, with the distribution for division one Bundesliga being the lowest out of the three. To explore this, the standard deviation of points after each round for the three trinomial models was calculated and is presented in Figure 3. As seen in Figure 3, the standard deviation of points for the trinomial distribution is much lower than the two observed divisions, with division one being greater than division two as the season progresses. The result can be explained by the constant match outcome probabilities that imply equal team strength, which was implemented with the trinomial simulation. This suggests that team strength should be taken into consideration when using simulation to calculate importance. We discuss this issue further in the Discussion section of this paper.

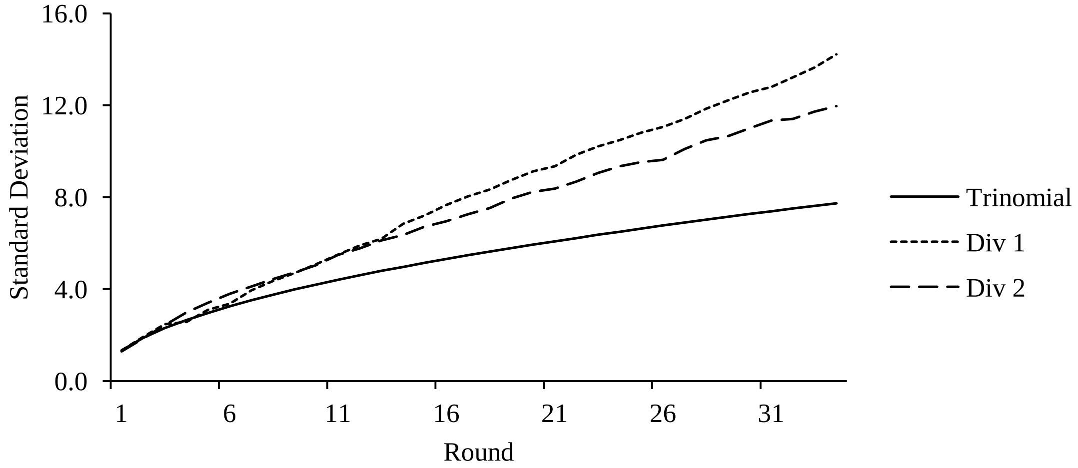



Figure 3: Standard Deviation of season points for simulated vs. observed seasons

## 4. DISCUSSION

One of the most common methods of measuring match importance in professional football is Monte Carlo simulation. This method simulates a season a number of times and determines which matches throughout a season are critical to a team's chances of winning the league championship (finish first place). A long runtime is usually required for simulation which is usually not ideal. In this paper, we presented a simple model to measure match importance, where a trinomial distribution calculates the conditional probability of a team finishing first given the outcome of the next match and their current season points total. Not only are the number of model inputs dramatically reduced, but computational runtime is reduced compared to the standard simulation model.

As mentioned in the methods section of this paper, the simulation model required a number of iterations to generate a season schedule, as well as further iterations to simulate season results and calculate the success probabilities. In this paper we elected to use 10,000 iterations for the season schedule and 100 iterations for the season results, where the latter was largely determined arbitrarily through trial and error. While it can be argued that a larger number of iterations are required to generate the success probability distributions, 100 iterations was found to provide a sufficient sample size to generate the required success probability distributions. Note here that 100 season iterations is not an optimised value, something that can be addressed in future research.

It was found that the average match importance for both divisions in observed seasons were slightly lower compared to the trinomial simulation. The standard deviation of the season points to date was also much higher in the observed seasons when compared to the simulation. The higher standard deviation for the observed data can be explained by the dominance of some teams, whose probability of winning the next contest would be far greater than the constant probabilities we have applied (e.g. Bayern Munich, who won 29 of 34 matches in the 2013/2014 Bundesliga season). Variations in team strength across a division would result in larger variations in season total points. This result suggests that team strength should be taken into account when measuring match importance, as both our trinomial simulation and observed models assume equal match outcome probabilities. The next step in building on this work will look to account for variations in team strength by adjusting these probabilities using a Bayesian approach. Such an approach could measure a team's recent form and as a result provide a better indication of their likelihood of winning their current match.

A key advantage of the trinomial distribution model over the simulation approach is that fewer inputs are required, due to the relaxation of match independence assumptions. While this can actually be a drawback, as clearly there is some dependence on other teams' results, the similarity of importance curves indicate that using the trinomial distribution model can produce reasonable estimates of match importance while reducing the computational runtime. This allows for one to easily calculate match importance during a live season while only evaluating the total points of the league leader and the current team of interest. While this paper has only focused on the importance of winning the league championship, we believe future work can assess other season outcomes, such as qualification for European tournaments or avoiding relegation.

While the results presented in this paper are promising, quantifying match importance in any professional sport remains a difficult task. Measuring match importance can be both subjective and objective; a fan would claim a match is important while the statistics may say otherwise. The best possible approach to actually quantifying match importance appears to be through simulation and, as we have discussed in this paper, a trinomial distribution approach produces similar results when assessing importance curves. If one were to include match importance as a variable in some predictive model, applying the trinomial distribution could yield a quick and reasonable result.

## 5. CONCLUSIONS

Quantifying match importance in professional football has often been completed using either a retrospective measure or complex computer simulation procedures. In this paper, we explored a probabilistic model that relaxes match outcome independence assumptions to counteract these drawbacks, finding that the importance distributions are similar to a Monte Carlo simulation approach. While the results are promising, further work is required to vary the match outcome probabilities to take into account team strength in football.

## References


Baimbridge, M., Cameron, S., & Dawson, P. (1996). Satellite television and the demand for football: A whole new ball game? *Scottish Journal of Political Economy, 43*(3), 317-333. doi:10.1111/j.1467-9485.1996.tb00848.x

Bedford, A., & Schembri, A. (2006). A Probability Based Approach for the Allocation of Player Draft Selections in Australian Rules Football. *Journal of Sports Science and Medicine*, *5*(4), 509-516.

Borland, J., & Lye, J. (1992). Attendance at Australian Rules football: A panel study. *Applied Economics, 24*(9), 1053-1058. doi:10.1080/00036849200000083

de Lorenzo, M., & Bedford, A. (2014). *An Alternative Draft System for the Allocation of Player Draft Selections in the National Basketball Association*. Paper presented at the Twelfth Australasian Conference on Mathematics and Computers in Sport (12 M&CS), Darwin, Australia.

Jennett, N. (1984). Attendances, uncertainty of outcome and policy in Scottish League Football. *Scottish Journal of Political Economy, 31*(2), 176-198. doi:10.1111/j.1467-9485.1984.tb00472.x

Lahvička, J. (2015). Using Monte Carlo Simulation to Calculate Match Importance: The Case of English Premier League. *Journal of Sports Economics*, *16*(4), 390-409. doi:10.1177/1527002513490172

Morley, B., & Thomas, D. (2005). An investigation of home advantage and other factors affecting outcomes in English one-day cricket matches. *Journal of Sports Sciences, 23*(3), 261-268. doi:10.1080/02640410410001730133

Morris, C. (1977). The most important point in tennis. In S. P. Ladany & R. E. Machol (Eds.), *Optimal Strategies in Sport* (pp. 131-140). Amsterdam: North-Holland Publishing Company.

Scarf, P. A., & Shi, X. (2008). The importance of a match in a tournament. *Computers & Operations Research, 35*(7), 2406-2418. doi:10.1016/j.cor.2006.11.005

Schilling, M. F. (1994). The Importance of a Game. *Mathematics Magazine, 67*(4), 282-288. doi:10.2307/2690849

# THE MATHEMATICS OF TENNIS

Tristan Barnett[ac], Alan Brown[b]

[a] *Strategic Games*
[b] *Affiliation not specified*
[c] *Corresponding author: strategicgames@hotmail.com*

## Abstract

The longest professional tennis match, in terms of both time and total games occurred at the first-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 book 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.

*Keywords*: Excel, recursion, distributions

# Statistical Modelling Tools for the Coaching of Elite Tennis Performance

Stephanie Kovalchik[a]

[a] *Victoria University and Tennis Australia, Melbourne, Australia*
*Correspondence: skovalchik@tennis.com.au*

## Abstract

Quantitative feedback is a fundamental tool in the coaching of sports performance. In elite tennis, the primary source of quantitative feedback about in-competition performance comes from computer-aided notational analysis systems that extract and summarise key performance events from video-recorded matches. As such systems have been in use for decades, large databases of performance outcomes have accrued. Although these data facilitate the statistical modelling of performance, the use of modelling in feedback systems for elite tennis is rare; single-match analyses and simple counts and percentage of events remain the norms for performance evaluation. In this talk, I present several novel statistical modelling approaches to contextualize the evaluation of performance and identify patterns of play. The developed methods include approaches for detecting outlying serve/return performance, measuring serve/return consistency, identifying back-to-the-wall and momentum effects, and assessing clutch ability. Visualizations are used to present the findings of each approach and make these insights more accessible to non-statisticians. Using specific player applications, I demonstrate how these model-based feedback tools can be used by coaches to identify areas where their player is in need of performance improvement as well as make the selection of video replays more precise.

*Keywords*: performance analysis, feedback, coaching