

Raising Interest in Statistics through Sporting Predictions on the Internet

Stephen R Clarke

*School of Mathematical Sciences
Swinburne University of Technology
PO Box 218, John St, Hawthorn, Australia
sclarke@swin.edu.au*

Researchers and teachers should consider using the web to publicize the results of their statistical projects to a wide audience. Such sites can create interest in statistics, and be used by teachers as a source of motivation, data or practical problems. For several years, Swinburne Sports Statistics has published on our web site www.swin.edu.au/sport, predictions of sporting events based on statistical modelling and simulations. The growth of sports betting has seen commercial application for the predictions and methods, and has led to the development of a subject 'Chance and Gaming', where probability is taught through applications in the gaming industry. The site received over 4,000 visits per week in 2002 and has raised interest in statistical methods in the general public and students. Some of the projects highlighted on our site, and the resultant publicity, will be discussed.

1. Introduction

Swinburne Sports Statistics has a 20-year history of forecasting sporting results through the media. With the introduction of the Internet, we began to place these predictions on our web site www.swin.edu.au/sport. Our aim was not just to provide subjective predictions, or ones based solely on historical data, but to demonstrate the uses of statistical modelling to 'add value' to data. Thus in addition to predicted winners, they usually contain estimates of probabilities of winning, or often of finishing in any position. In addition to creating interest in the general public, with over 4000 visits per week in 2002, the predictions have often resulted in media publicity. Here we discuss some of the topics we have highlighted on our web site.

2. Australian Rules Football

Australian rules football is the major winter spectator sport in the southern states of Australia. The sport receives huge media publicity, and the author has published computer predictions of Australian rules football for over 20 years (Clarke, 1988; 1993). While media outlets traditionally published only the winner predicted by their resident experts, the computer predictions always went much further. Based on statistical modelling and simulations, they contained predicted margins, chance of winning, chance of any possible finishing position at the end of the year etc. Since the media was originally only interested in the winner, the WWW became a vehicle for publishing the extra predictions direct to the

public. We began to put the predictions, along with comparisons with other media experts, on our web site. Since the single computer prediction was always compared with the best of several media commentators, I encouraged other staff or graduate students to also create computer predictions. This has resulted in various computer models, using different data and assumptions, and we currently publish three different models. Ironically, the growth of sports betting means that the original computer predictions, with their objective estimates of chances of winning, now have a commercial value, and are sold to punters via a subscription service. This has meant they can no longer be provided free, and we only publish predicted winners on the web for this model.

The Australian public is addicted to office football tipping competitions, which require entrants to enter their predicted winners each week. Entrants usually include the CEO down to the tea lady, the rabid football follower down to the otherwise disinterested onlooker. Expert and novice alike all like to get some extra impartial advice from the computer. Over 1500 people per week check our web site for the computer's predictions before making their own selection. Many just copy the computer's tips. In much the same way as pollsters use election predictions to demonstrate the accuracy of their sampling methods, the computer tips demonstrate to the public the worth of statistical methods of forecasting. The following quotes are typical of the email received in reference to the site.

“Many thanks for your tips each week...Thanks to you I won the tipping comp at my office. On those occasions I thought I knew better I usually bombed out !!” Mike

“I religiously look up your wonderfully helpful site each week - in the hope of improving my tipping ability.” Cathy

“I notice you have not updated your football site. I unashamedly on bended knees implore you to update your site immediately. I'm winning the office tipping due entirely to your wonderful system. This is my only chance of getting anywhere in the tipping polls. You see, I am a StKilda supporter, floating alone in a sea of Weagles supporters, and if I rely on my own ability to pick winners emotion gets in the way” Phillip

The tips have proved so successful, we have diversified into other sports – we currently have two different weekly tips of Rugby League and one of Netball. In fact these are all supplied in part by non-Swinburne personnel. I often receive emails from the public expressing interest in the techniques used, and in applying them to other sports.

3. Tennis

With a view to predicting the 1998 Wimbledon championship, David Dyte (then our Sports Statistician) and I developed a logistic regression model which used the ATP rating points of the two players to estimate their chances of winning a set (Clarke & Dyte, 2000). Once the draw was made, it was relatively simple to simulate the whole tournament. A random number was generated to decide the winner of each set, and the match result tracked. The simulated winners were then advanced to the next round according to the actual draw,

until a final winner determined. 10 000 runs were used to generate estimates of the probabilities of each player winning the tournament or making the semi-final. The results were updated on our site at the end of each day. Each day, the players with the highest estimated chances of winning were published. The method was also applied to the 1998 US Open and 1999 Australian Open, and later majors.

One advantage of simulation is that the probability of virtually any event of interest can be estimated. While the main interest here was the probability of each player progressing to a certain stage, the chance of compound events is easily produced. The chance of two given players meeting, the chance the final will contain an Australian, the chance one player will progress further than another, can all be calculated. Because of limited space, only the chances of the main players could be tabulated. With an international event such as this, there may be wide international interest in a player of little interest to Australians. Rather than just provide static tables, Myles Harding, a professor in our school with expertise in computing, designed a java applet that allowed any single player to be selected. The applet then gave that player's chance of reaching each round. Alternatively, two players of interest could be selected, and the applet gave the round they would meet and the chances they would meet.

Such sites require a concerted effort over an extended period of 2 weeks. With international events, this effort is often required outside normal working hours. The individuality of such modelling also makes it difficult to continue when someone instrumental to the project leaves. Tennis tips have been provided by several personal at Swinburne, but the original model for simulating the tournament is no longer used. We currently provide predictions for matches at tennis majors through Tristan Barnett, a PhD student, who provides predictions of not only the winner, but expected score lines and length of matches (Barnett & Clarke, 2002). However this is not extended to chances of winning the tournament. We usually receive publicity for our predictions. For example, in 2003 Tristan had two newspaper articles written about his work and conducted a couple of radio interviews.

4. Brownlow medal

The Brownlow medal is a prize awarded to the best and fairest player in the Australian Football League on the basis of 3,2,1 votes given in each game by the umpires. The votes are kept secret until the end of season, when they are counted round by round in what has become a major television event. Betting takes place throughout the year, and even throughout the night as counting progresses. In 2000 we decided to produce a prediction model for the umpire votes based on various playing statistics (Bailey & Clarke, 2002). The analysis produces expected votes, predicted finishing order and chance of winning. Myles Harding implemented an Internet system that allowed our predictions of final votes to be updated live during the count, with predicted votes being replaced by actual votes. We have run this process live each year over the internet, but such a system would also be suitable for delivery over the telephone lines. This product has been provided free over our web site, but the predictions were published for a fee in 2001 (Bailey, 2001). We are hopeful the television station broadcasting the count might take it up in 2003. At the moment, they utilise a leader board based on votes counted, but this may bear little relation to the final votes.

Michael Bailey, then our sports statistician, did the original regression study to predict the number of votes the umpires gave each player based on the performance statistics they obtained during the match. He discovered that some players with certain physical characteristics tended to have large residuals and obtained more votes than predicted, so he built this into his model. When we published our predictions, one newspaper article highlighted the fact that the umpires favoured players with blond hair or dark skin (Anon, 2000a). The morning after the actual count, I was feeling somewhat sheepish as our predicted winners had come second and third, and we had suggested the actual winner would come about tenth. But a newspaper and radio story trumpeted that Swinburne had got it right, "Swinburne University's sports statistics department identified last night's Brownlow medal winner before a vote was counted" (Anon, 2000b). The justification for this statement was simply that the winner had blond hair. We had no input to this article - it had been generated by a journalist from the earlier newspaper article.

5. Grand Prix

Early in 2000 I was approached by the media people associated with the Grand Prix racing in Australia – 500 cc motor cycling at Phillip Island and F1 racing at Albert Park. They had seen the publicity generated by our successful predictions in other sports, and wanted us to try to predict their events. Jonathan Lowe, at that time a PhD student, took up the challenge. We devised a simulation based on estimated times behind the winner, and chances of not finishing the race. Using 10,000 runs the system was trialed with some success for the 2001 motor cycle Grand Prix at Phillip Island. The simulation predicted the chance of each rider finishing in any position in the race, but also the World Championship. We received media coverage of the predictions (Anon, 2001), and the success of the system was assured when the two favourites came first and second in the race.

Following the success in motorcycles, we applied similar methods to the 2002 F1 Grand Prix. This produced special problems. While the Phillip Island race was the last race of the motorcycle season, the Albert Park race is the first of the F1 season. With a possible change of drivers over the break, we had to introduce a car effect and a driver effect. The output consisted of four tables, the chances of all possible finishing positions for both drivers and manufacturers in both the Melbourne race and the World championships. The predictions received extensive coverage, including mentions in British magazines (Anon, 2002) and a German newspaper. In addition to publishing on our site, for the last two years, the Swinburne predictions have been supplied as part of the media handout at the Melbourne launch. This has allowed Jonathan and myself photo opportunities with the F1 cars and the pit girls.

6. Cricket

When David Dyte joined us as a Sports Statistician in 1998 he placed on the site a cricket game he had previously developed with others. The game was not high tech in appearance, but had some heavy mathematics behind the scenes (Dyte, 1998). The game allows players to select two teams, and then play the game by selecting batsman and bowlers. The game produces over by over scores, and reproduces a scorecard. The statistical modelling

behind the game ensures that over many games, batsmen and bowlers reproduce their test averages. The game continues to receive many hits, and periodic requests to update the current player averages or update the lists with new players. The success of the game shows people are interested not just in flashy front ends, but also in valid statistical modelling behind the process.

7. Further developments

The development of our web site has gone hand in hand with developments in other areas. The growth of sports betting, particularly over the Internet, has seen commercial application for the predictions and methods used on our site, and in worldwide interest in odds predictions, even for local events. Our site now has discussions on the relationship between probability and gaming, and the use of various wagering systems. This has led to the development of a subject 'Chance and Gaming', where we use Excel to apply probability concepts to the gaming industry. The aims of the subject include introducing students to elementary probability and distribution theory, giving students an understanding of the gaming and wagering industry, and illustrating the use of Microsoft Excel as a tool for solving a range of gaming problems. We have also developed a subject Modeling in Sport, where many of the techniques used in the site are taught.

Many visitors to our site follow links from our Sports Statistics site to the school of mathematics site, or our statistics@swinburne site www.swin.edu.au/statistics, which should assist with enrolments. This in turn creates a pool of interested people, who then link their material to the site, or are useful in obtaining further publicity. For example, Don Forbes was an interstate student originally attracted to our site by our sporting predictions. Don's aim in life is to become a bookmaker. He enrolled for the Graduate Diploma and did the Chance and Gaming subject, and is now a PhD student doing a study of Dynamic prediction in AFL football. Through his industry partner, he now regularly provides the statistical analysis and even the text that is the basis of a regular weekly article the Australian Financial Review. In 2002 I was asked by a TV program to provide a Swinburne tipster for our famous horse race, The Melbourne Cup. This race stops the nation once a year. Although it had nothing to do with his project, knowing his interest in the horses I suggested Don, with the result he was part of a segment on national television. Prior to his appearance, the media people at Swinburne gave Don some coaching, where the main advice was to keep it simple – don't get too technical. However when the interview took place, the producer actually asked Don to be as technical as possible, as they wanted to contrast his methods with that of an astrologer and a dog. Don obliged by talking regression, and the interviewer then pretended to sleep. However we believe humour is one way of raising interest, and the segment provided an amusing example of using statistics.

8. Conclusion

Our Web site received over 4,000 visits per week in 2002. This raises interest in statistical methods in the general public and in potential and current students, and has resulted in many other media and consulting opportunities. Many of the projects originated purely to

feature on our site, but have ultimately resulted in refereed publications. Researchers and teachers should consider using the web to publicize the results of their statistical projects to a wide audience. Such sites can create interest in statistics, and be used by teachers as a source of motivation, data or practical problems. Two such sites are www.amstat.org/sections/sis which concentrates on statistics in sport, and www.statistics.com with links to statistical articles, data, problems and other information suitable for statistics students.

The web is just one way of raising interest in statistics. Clarke (2002) discusses creating interest and publicity in mathematical and statistical methods through the use of other methods such as press releases, with a view to publishing in the popular press and technical journals.

REFERENCES

- Anon (2000a). Hairs to all those Brownlow hopefuls. *The Age*. August 23, Melbourne: p 3 (Sport section).
- Anon (2000b). Being blond gets votes. *Herald Sun*. August 29, Melbourne.
- Anon (2001). Computer odds favour the doc. *Australian Motorcycle News*: Oct 12, p 14
- Anon (2002). Computer simulation - championship predictions. *F1 Magazine: Season Programme 2002*: 18-20.
- Bailey, M. (2001). Who'll win the Brownlow: Permutations, combinations, probabilities and guesswork. *AFL Record*. 1st preliminary final: 80-81.
- Bailey, M. and S. R. Clarke (2002). Predicting the Brownlow medal winner. *Proceedings of the sixth Australian conference on mathematics and computers in sport*. G. Cohen and T. Langtry (Eds.). Gold Coast, University of Technology, Sydney: 56-62.
- Barnett, T. J. and S. R. Clarke (2002). Using Microsoft Excel to model a tennis match. *Proceedings of the sixth Australian conference on mathematics and computers in sport*. G. Cohen and T. Langtry (Eds.). Sydney, University of Technology Sydney: 63-68.
- Clarke, S. R. (1988). Tinhead the tipster. *OR Insight* 1(1): 18-20.
- Clarke, S. R. (1993). Computer forecasting of Australian Rules football for a daily newspaper. *Journal of the Operational Research Society* 44(8): 753-759.
- Clarke, S. R. and D. Dyte (2000). Using official ratings to simulate major tennis tournaments. *International Transactions in Operational Research* 7: 585-594.
- Clarke, S. R. (2002). Publicising OR projects in the media. *OR Insight* 15(3): 4-14.
- Dyte, D. (1998). Constructing a plausible test cricket simulation using available real world data. *Fourth Conference on Mathematics and computers in sport*. N. de Mestre and K. Kumar (Eds.). Gold Coast, Qld, Bond University: 153-159