

Successful applications of statistical modeling to betting markets

Stephen R. Clarke*, **Michael Bailey**** and **Stefan Yelas*****

Faculty of Life and Social Sciences, Swinburne University of Technology, PO Box 218
Hawthorn, Victoria, 3122 AUSTRALIA

With a strategic theme of entrepreneurship, Swinburne University encourages innovation and commercialisation of research results. Swinburne Sports Statistics initially embraced this theme by selling computer predictions of Australian rules football to the media, and since 1998 predictions for a range of sporting events were placed on our web site. The expansion of sports betting created a market for sports prediction expertise, and we discuss profitable attempts at demonstrating the inefficiency of football, cricket and rugby betting markets. The head to head nature of most sports betting results in low margins for bookmakers, and the lengthy time many events take to complete also results in higher risk and lower turnover. To overcome these drawbacks, 'Betting in Running' allows betting on events which occur within a sporting contest, such as the number of runs in an over of cricket, or the point score in a single game of tennis. Sportsbet21, a Swinburne startup company, provides to bookmakers computer generated odds driven by a statistical model. Models have been developed for both cricket and tennis, and operated successfully by Ladbrokes in the UK for some time. The product has achieved profit targets on growing turnover, and demonstrated the robustness of the mathematical models.

Keywords: AFL football, cricket, rugby, inefficiency, gambling

Introduction

Swinburne Sports Statistics has been in the business of predicting sporting results for over 25 years. Our forays into sports prediction started as a student project in 1980 to predict the now Australian Football League (AFL) finalists. This resulted in a computer prediction program (Clarke, 1993), the results of which have been published in various media outlets ever since.

* sclarke@swin.edu.au

** michael.bailey@med.monash.edu.au

*** stefan.yelas@beaton.com.au

With the appointment of a research assistant in 1997 we began to investigate and publish predictions of other sports, both in the professional literature and on our web site. These predictions were all based on mathematical or statistical modeling performed by staff or students, or sometimes by staff at other institutions. Sports predicted included AFL football, cricket, soccer tennis, F1 motor racing, Grand Prix motorcycle racing, horse racing, netball, rugby, rugby league, baseball, basketball, beach volleyball and The Olympic Games.

While media experts generally only predict winners of sports events, our aim was to demonstrate the power of modeling to add value to data. Our predictions usually predicted not only the winner, but other statistics of interest, such as margin of victory, chance of winning, and chance of any possible finishing order. For many years in Australia, betting was only allowed on horse and dog racing. However, as Governments relaxed this restriction and betting began on a range of sports, it became clear our expertise in prediction might be put to profitable use in the gaming industry. Our unit began to attract students with an interest in using their mathematical skills to make money from the bookmakers, or in academic parlance, testing the efficiency of the growing sports betting markets.

This paper discusses some successful applications of mathematical and statistical modeling to betting markets. We first discuss some gambling principles, then discuss profitable attempts at demonstrating the inefficiency of football, cricket and rugby betting markets. Finally we discuss the development of Sportsbet21, a Swinburne startup company which offers computer model generated odds on events within a sporting contest. This system is designed to overcome some deficiencies of sports betting.

1. Gambling principles

Haigh (1999) gives a good introduction to some of the mathematics behind betting. The most important parameter is the

house percentage. This is the expected return to the operator as a percentage of turnover, and is effectively the expected cost to the punter of participating. For most casino games it can be calculated exactly and is usually independent of the punters' choices. For example most lotteries work on about 40%, poker machines 10-15% and roulette 2.7%. Totalisator betting systems for thoroughbred, dog and harness racing usually extract an operator profit in the range 15-20%. Games of strategy such as Blackjack depend on the choices of the player, and it has been shown that for a player using optimal strategy and counting cards (which is illegal) the house percentage is negative. For fixed price horse racing and other sports events, the house percentage is a theoretical figure that is only achieved if a bookmaker balances his book. For example, a typical horse race might have 8 runners paying \$2.50, \$4, \$5, \$10, \$10, \$20, \$20, and \$50 for a \$1 bet. The bookmaker attempts to get an amount bet on each horse proportional to $100/\text{price}$. If successful, he has accepted bets totaling $100/2.5 + 100/4 + \dots + 100/50 = \117 , but pays out \$100 no matter which horse wins. His percentage profit is $100 * 17/117 = 14.5\%$. Note that if each horse is treated independently, the bookmaker can take a much bigger cut from the long shots than from the favourites. To take an extreme example, a runner whose true price is \$1000 (probability of winning = .001) could be priced at \$500, and the expected return to the punter is 50c. But a very short priced favourite whose true price is \$1.05 (probability of winning = 0.95), even if priced at \$1.01 returns on average 96% to the punter. It is this characteristic which makes sports betting more attractive than traditional racing. Because many events are head to head contests with only two outcomes, many sports books have a house percentage of 5% or even 2%. The astute punter does not have to perform much better than average to turn a profit.

A second important characteristic is the speed of the game. A lottery punter loses 40% once a week, a punter on the horses averages 15% loss every 45 minutes, a poker machine takes its 12% every 10 seconds. Another contributor to profits is churn – the effect of reinvesting winnings. A major lottery winner rarely spends all his winnings on more tickets, but a racing punter will reinvest a proportion of winnings on the next race. Poker machines automatically add winnings to the stake a sucker is playing, so apart from major prizes they replay all their wins.

Since many sports events are one sided, a bookmaker has trouble balancing his books. If Australia plays Tonga at rugby, Australia might be \$1.01 and Tonga \$15, but at those prices there would be very little interest. Line betting allows the bookmaker to quote a winning margin (say Australia by 29.5 points) and a punter then bets on Australia to win by 30 or more, or win by 29 or less. These two outcomes might pay \$1.90 each, for a house percentage of 5%, but competition often results in payouts for the line as high as \$1.95 for a house percentage as low as 2.5%.

The aim in profitable betting using models is not to bet on whom the model predicts will win, but on value outcomes with a positive expectation – i.e. those outcomes whose estimated probability of winning is greater than the inverse of the price. Thus if a team has only a 30% chance of winning, but their price is \$4, our expected return is $30\% \times \$4 = \1.20 for a 20% gain or overlay. In an area like football betting, where followers often bet with their hearts, and a few highly publicised commentators or events can influence public opinion, prices may not reflect the true probabilities. A mathematical model can be used as an objective guide to a team's chances and hence to evaluate bet value. Now of

course the computer is not exactly correct. Because of the error associated with any model prediction, most punters allow for a margin of error. If you bet on small overlays, you will have lots of bets at what should be reasonable odds. If you bet on large overlays, you will have a few bets at very good odds.

If you have a system where you are making bets with a positive expectation, then you will make a profit in the long run. But you may still go broke in the short run (that is why casinos have house limits). You need to balance the desire to bet large (as you make an average percentage gain on each bet), and the increased chance large bets will mean a run of bad luck will use all your capital. Profitable betting is as much about suitable money management to avoid going broke in the short term as finding bets that will turn a profit in the long term. In general, the bigger the overlay, the larger the bet. On the other hand, the smaller the chance of winning the smaller the bet. Thus it is much riskier betting on a 1 in 10 chance showing a 15% overlay, than a 50-50 chance paying a 15% overlay.

Kelly (1956) showed that the percentage growth rate in your pool is optimized if you bet a percentage of your pool each bet equal to the expected profit/maximum profit, i.e. the percentage overlay / (price - 1). Thus for a 15% overlay paying \$3 we should bet $15\% / (3-1) = 7.5\%$ of our pool, but for a 15% overlay paying \$6 we should bet $15\% / (6-1) = 3\%$ of our pool. The Kelly system relies on some assumptions that don't really apply in many cases. When applied in a real situation, the Kelly system is extremely volatile, and carries a large risk that your pool will disappear. In a long sequence of bets you will always have a long run of losing bets, and this *will* wipe out your pool if you are using Kelly. This risk *never* diminishes, as bets are proportional to pool size. As your pool builds up, your risk of losing everything is the same as at the beginning. To reduce this risk, many punters prefer to use a half Kelly, one third Kelly or some other fraction. Some might also place an upper limit on the bet size, such as 1%, 5%, 10% or 20% of the pool. Clearly a chosen betting strategy depends on the psychological makeup of the punter and a punter's goals. One punter may prefer small steady returns on a large capital base with little risk of losing all capital. Another may be happy to go for broke and put \$500 into the wild ride they will get from the full Kelly, with a large chance of losing all in the hope of scoring big, rather than buy a system 20 Tattsлото entry. Only the punter knows his risk taking makeup, financial situation, and betting goals.

2. AFL football

Swinburne's AFL computer tips have been published in various media outlets since 1981. Although in the main the media only published predicted winners, the program predicts not only the winner of an upcoming match, but also the expected margin in points, the chance of each team winning, and via a simulation of the remainder of the year, the chance of each team finishing the home and away series in any position. With the growth in sports betting this information has become of interest to punters. From the beginning, we had envisaged the possibility of marketing the predictions direct to interested parties via a mail subscription, but lacked the marketing skills to implement. At one time we had a proposal to market through a telephone service, but there was little incentive to proceed. With the financial policies in place at the time, the main motivation for marketing the tips was the publicity the school and the University received.

The program was ahead of its time in publishing margins and chances of winning long before sports betting on these outcomes was legal. Betting is now allowed on Australian rules football. In addition to head to head betting on the winner, various bets can be made on the margin of victory and on teams making the finals or winning the premiership. Various analyses performed in the past suggested it was possible to exploit market inefficiencies using the computer's predictions. In 2001, a research student obtained the past odds on each team winning and simulated various betting strategies based on the computer's tips for 3 years from 1998 to mid 2001. These showed a substantial profit for some strategies. Yelas (2003) analysed data from 1998 to 2002 and showed that all betting strategies based on a safety margin of 5% overlay or more showed a profit over the 5 year period. In 2002, following an approach from Ozmium Ltd the program was marketed to punters. Ozmium were already marketing horse racing tips through their web site <http://www.smartgambler.com.au/> and were keen to diversify. The tips were emailed to the list of subscribers each week, along with a

spreadsheet showing a bookmaker's prices, overlays, and bet sizes according to the Kelly formula. The punters placed their own bets, based on the programs recommendations, their own information and betting strategies. A discussion list was set up, and subscribers discussed the tips, their own systems, the pros and cons of various bookmakers, good deals etc. In line with our recommendation, it was quite clear that most subscribers were not using the computer's predictions as a black box, but as a starting point to which they added their own opinions. It was also clear that they all had their own betting systems. While some bet on all predictions which showed some set percentage overlay, others required a certain probability of winning etc. Table 1 shows the returns achieved in 2004 by betting a constant \$10 with TAB Sportsbet on teams showing a given overlay.

Note that potential gains are much greater than is apparent from the table, and the ROI figure should not be confused with an annual rate. By adding winnings to the betting pool, and betting a percentage of a growing pool, a punter might achieve this growth every month or week. For example, using Kelly recommended

Table 1: Returns from a constant bet of \$10 with TAB Sportsbet 2004 on all teams with given overlay

Overlay	Number of bets	Number of winning bets	Amount bet \$	Amount won \$	Profit \$	ROI
0%	116	80	1,160	1,332	172	14.8%
2%	96	66	960	1,113	153	15.9%
4%	87	59	870	1,006	136	15.7%
6%	72	48	720	859	139	19.3%
8%	52	40	520	727	207	39.8%
10%	37	27	370	488	118	31.9%
12%	26	21	260	399	139	53.3%
14%	18	13	180	278	98	54.2%
16%	13	10	130	221	91	69.6%
18%	12	9	120	207	87	72.1%
20%	10	7	100	174	74	73.5%

Table 2: Returns from betting the Kelly percentage of a variable pool with a \$100 start with TAB Sportsbet 2004 on all teams with given overlay

Overlay	Amount bet \$	Amount won \$	ROI	Final Pool \$	Min Pool \$	Max Pool \$
0%	14,791	16,764	13.3%	2,073	87	3,076
2%	13,532	15,352	13.4%	1,920	87	2,885
4%	12,241	13,907	13.6%	1,765	87	2,653
6%	9,611	11,153	16.0%	1,642	87	2,468
8%	12,926	15,402	19.1%	2,575	100	3,870
10%	5,424	6,741	24.3%	1,416	100	2,128
12%	3,772	4,736	25.6%	1,064	100	1,599
14%	1,079	1,383	28.2%	404	100	608
16%	915	1,268	38.6%	453	100	681
18%	739	1,024	38.5%	385	100	578
20%	454	624	37.3%	269	100	405

Table 3: Results of fixed bet size \$10 on the line at globalsportsbook \$1.95 odds

Safety margin	Number of bets	Number of winning bets	Amount bet \$	Amount won \$	Profit \$	ROI
0	184	105	1,840	2,048	208	11.3%
1	163	95	1,630	1,853	223	13.7%
2	148	89	1,480	1,736	256	17.3%
4	120	77	1,200	1,502	302	25.1%
6	92	60	920	1,170	250	27.2%
8	72	48	720	936	216	30.0%
10	53	35	530	683	153	28.8%
12	41	27	410	527	117	28.4%
14	30	20	300	390	90	30.0%
16	19	12	190	234	44	23.2%
18	13	10	130	195	65	50.0%
20	11	8	110	156	46	41.8%
22	9	6	90	117	27	30.0%
25	3	3	30	59	29	95.0%

sized bets, the 87 bets placed by the punter above betting on all overlays of 4% or more, would have turned a \$100 starting pool into \$1765 by the end of the season. Because of the growing pool and bet sizes, the punter would have bet a total of \$12,241 for a return of \$13,906, and at one stage the pool would have reached \$2653. Results for all overlays are shown in Table 2. Clearly a profitable year like this would allow a punter using this risky and volatile betting system to be ahead even if their starting pool was wiped out in several subsequent years.

Because the computer does not use team personnel for its predictions, bets can be made as soon as the previous weeks matches are complete. The above tables are obtained using odds posted on Monday. During the week, the odds tend to change due to the weight of money and team selections. Because the bookmakers become more certain of public opinion and the true odds, the odds tend to tighten with a reduced bookmaker's margin. However these competing forces have little effect on the profitability of the computer's estimates. Clarke and Clarke (2006) show constant size bets on all given overlays result in return on investment ranging from 22% to 47%. Because of the lower bookmaker's margins, more bets are made using Friday's odds. The returns on investment are generally higher for small overlays (perhaps due to the lower margins) and lower for the higher overlays (perhaps due to the public taking advantage of the more serious mistakes in setting the early odds).

Since the computer selects margins better than winners relative to humans, one might expect betting on margins to be a better proposition than head to head. Lines, which allow betting at the margin for a return of \$1.95, are quoted each Friday by Globalsportsbet. These have a low bookies margin of only 2.5%, and given the computer's superiority in predicting margins should be a prime opportunity. The lines were obtained from each Friday for each week except for round 3 and one match in Round 13. For these missing values, an approximate line based on the Monday TAB sportsbet odds was used.

For line betting, the computer gives an expected margin, so either side of this margin would be expected to occur with 50% chance. A safety margin can be built in by requiring the line to differ from the computer's prediction by at least a given figure. However betting on the line with no safety margin in 2004 would have resulted in 105 winning bets out of 184, for a return on investment of 11.3%. Details for this and other safety margins are shown in Table 3.

While 2004 was a stellar year, results in 2005 and 2006 have been less consistent, and highly dependent on the betting strategy used. However it is clear that inefficiencies in AFL betting markets can be exploited even by simple models. It is worth emphasizing that the above model uses only the team names, grounds and final scores of previous matches as input data to produce prediction variables of team form and home advantage. More comprehensive models yield even better results. Using data from 100 years of football, Bailey and Clarke (2004a) developed 3 prediction models and tested their efficiency on a holdout sample of 7 years of data from 1997 to 2003 inclusive (1786 matches). The model using the same form as above showed a return on investment of 1.3%. A second which also included extra predictors of very recent form, distance traveled to matches and experience of the team at the ground, yielded a return of investment of 10.1%. The best performance of 15.1% was obtained by the model which allowed for team selection, by combining the above measures for each individual player selected.

3. Cricket

In 2000, Michael Bailey took up the position of research assistant with Swinburne Sports Statistics, and subsequently enrolled as a PhD student. Michael's research was generally driven by a desire to exploit inefficiencies in sports betting markets, and he successfully modeled various outcomes in AFL football and cricket (Bailey, 2000; 2005, Bailey & Clarke, 2002; 2004a; 2006). Bailey & Clarke (2004b) describes a typical analysis of betting on

cricket. A One-day International cricket match (ODI) takes a day to complete, and since simple Head to Head betting is too restrictive, bookmakers offer other bets relevant to the game. One such is betting on which of two nominated batsmen will have the higher score. The approach taken was to first estimate the expected score of each batsman. 1936 ODIs comprising 33,606 individual batting performances were analysed to find the factors which affect batting performance. Factors such as career batting average, time of day of innings, innings order, home country, quality of opposition, number of games played, team, and past performance against the particular opposition were all found to be significant in a multivariate model. Several models for predicting scores were tested, with a log linear model proving the best predictor on a holdout sample of every fourth match.

Converting expected scores to probabilities is simplified by assuming scores are geometric. Elderton (1927) used scores of batsmen to illustrate the exponential distribution (the continuous equivalent of the geometric), and Wood (1945) used the geometric distribution of scores to investigate consistency. Kimber & Hansford (1993) have confirmed this is approximately true, with small departures around zero and for very large scores. Let

$$\Pr(\text{Score} = x) = p(1-p)^x; x = 0, 1, 2, \dots$$

p can be thought of as the probability of being dismissed before scoring another run. The distribution has a mean of $\mu = (1-p)/p$ and is fitted using $p = 1/(1+\bar{x})$. It is easily shown the $\Pr(\text{Score} \geq x) = (1-p)^x$, so for two batsmen we can calculate the chance one will outscore the other by summing the conditional probabilities the first will outscore the second, for each possible score of the first. This gives the rather elegant result that for two batsmen with expected mean scores m and n respectively,

$$\begin{aligned} \text{The probability the first will outscore the second} \\ = m / (m + n + 1) \end{aligned}$$

$$\text{The probability of equal scores} = 1 / (m + n + 1)$$

$$\begin{aligned} \text{The probability the second will outscore the first} \\ = n / (m + n + 1) \end{aligned}$$

From the 50 completed matches in the 2003 World Cup, prices were collected from three Internet sports bookmakers for 263 head to head batting match-ups, with the bookmaker's percentage approximately 7%. The above equations were used to convert to odds the predicted scores of four different multivariate models and 6 univariate models, and the efficiency tested using a Kelly betting system on a fixed pool of \$1000. The univariate models showed no ability to exploit inefficiencies, with return of investments ranging from 2.4% to -34.6%. However each of the multivariate models yielded return on investments of at least 20%. The best performer was a multivariate log linear model, which would have produced 107 bets for a total outlay of \$7480 and a profit of \$2720 – a return on investment of 36.3%. Of course such results are tempered by the fact the analysis was performed after the World Cup. Such post event analysis is of little help to the punter. However a statistically significant relationship did exist between the mean absolute error of a model's pre World Cup predictions and the return on investment. This suggests that better forecasts prior to an event are a pointer to profitable models. In fact 99 bets were actually placed by Michael during the World Cup for a return on investment of 16%.

4. Rugby

In 2003 we were keen to produce predictions for the upcoming World Cup in Rugby, and a graduate student Stefan Yelas agreed to take it on as a project in the subject Sports Performance Modeling. The aim was to produce individual match predictions, and also via a simulation, predictions for the tournament as a whole. We wished to build a forecasting model to predict not only the winner of each game of the 2003 Rugby World Cup, but also each team's chance of a place finish (first to eighth) at any given time in the tournament. Several models were to be explored based on the exponential smoothing method used by Clarke (1993) to predict the outcome of AFL matches. The best performing model would be used for publishing predictions on the Swinburne Sports Statistics website.

Three models, with varying degrees of complexity, were eventually built to predict the signed margin. Each model used team ratings and a common home advantage. An exponential smoothing constant was used to update team ratings depending on how the predicted margin differed from the actual margin. The models used exactly the same input data and only varied by either the prediction equation or the measure of error used to update the ratings.

The first model used standard exponential smoothing of team ratings to produce a margin prediction. The ratings were updated using simple exponential smoothing of the actual error. Thus

$$\text{Predicted margin} =$$

$$\text{Team 1 rating} + \text{Home advantage} - \text{Team 2 rating}$$

$$\text{Updated Team 1 rating} =$$

$$\text{Team 1 rating} + \alpha (\text{predicted margin} - \text{Actual margin})$$

$$\text{Updated Team 2 rating} =$$

$$\text{Team 2 rating} - \alpha (\text{predicted margin} - \text{Actual margin})$$

The second model used the square root of the predicted and actual margins in the updating equations to reduce the relative importance of large errors and magnify the relative effect of errors near the win/loss boundary. The third model, used an attack and defence rating for each team similar to one used by Morton (2002).

The models were fitted using all 566 matches between any of the 20 competing teams from 1996 to the start of the tournament. The methods were compared on the optimal value of the percentage of games correctly predicted, the standard deviation and the average absolute value of the errors over the last 50 predictions as well as all the data. In fact there was little difference in predictive power, and the simplest first method performed as well as any, correctly predicting 86% of winners in the last 50 games with an average absolute error of 11.8 points. It was decided to use this method for the live predictions, with a common home advantage of 5 points (to be enjoyed only by Australia in the tournament). The normal distribution was used to convert margins into probabilities for individual matches, and a simulation used to produce estimates of tournament performance. At regular intervals predictions were placed on our web site comprising the winner, margins and chance of winning for each match, along with chances of each team winning or coming second in their pool, and finishing in each of the first four positions in the tournament.

It should be said that we did not really expect the model to perform very well. The ability of teams in the tournament covers a huge range, and many matches were very one sided. For example the biggest margin in the tournament occurred when

Australia beat Namibia by 142 points. Most pundits would believe that the margin of victory when powerful rugby nations such as Australia, England and New Zealand play such minnows tells you little about their performance against each other. In fact the model performed far better than expected in a competition with many one-sided matches. It correctly predicted 46 of the 48 matches. One match it got wrong was the final, where it predicted Australia to win by 1 point. In fact England won in extra time, after the score was level at full time.

An obvious application for the model is sports betting and as the pool games are often one sided the logical application is margin betting. After the first week of the competition (once the model had been seen to work) a betting syndicate consisting of two of the current authors and a third colleague placed the equivalent of a \$75 pool with website www.betfair.com. Betfair is a betting market where members can lay or take bets. Bet Fair only takes a percentage on winning bets. Using the predicted probabilities and margins, bets were placed on games where a 20% or higher overlay was apparent. A fraction of the pool of 40% of the probability of the predicted margin range was bet. At the end of the competition the pool was nearly \$1000. Note that even the semi final and final matches, the only two matches in which the model predicted the incorrect winner, resulted in a significant contribution to this profit. In both cases the model suggested Australia would do better than the line indicated. As these matches were late in the tournament, the bets were a reasonable proportion of what was, by then, a relatively sizeable pool. Further details can be found in Yelas & Clarke (2004)

The syndicate continued to operate after the World Cup, applying the same methods to a range of sports including Australian rules football, Super 14, New Zealand, European and International rugby, and American football. The volatility of using a proportion of the pool as the betting unit is demonstrated by the pool at one stage reaching \$15000 only to plummet to \$2000 in a couple of weeks. However by early 2006 the pool was approximately \$5000 having paid out dividends totaling \$10,860 – an excellent return in 3 years for an investment of \$75, made even more surprising by the relatively simple methods used.

Cricket: ICC Champions Trophy
Australia vs West Indies
 LIVE NOW

SCORE
 Cricket Over 31: Cricket Runs 3
 Cricket Over 30: Cricket Runs 5
 Cricket Over 29: Cricket Runs 10

Scorecard

Match betting
 Australia 1.22 West Indies 4.00

Next batsman out
 Selection Odds
 5th - wicket - R Morton 1.61
 5th - wicket - B Lara 2.20

Next over - 1st innings
 Innings 1, Over 33
 Selection Odds
 0-1 Runs 5.00
 2-3 Runs 3.00
 4-5 Runs 3.00
 6-7 Runs 5.00
 8-9 Runs 9.00
 10-11 Runs 15.28

Today
 Cricket
 Australia vs West Indies
 Tennis
 F Gonzalez vs P Srichaphan
 J Blake vs K Wliegen
 I Ljubicic vs A Murray
 A Mauresmo vs S Stosur
 R Soderling vs J.C. Ferrero
 D Hantuchova vs A Sugiyama
 R Nadal vs M Fish
 M Safin vs M Baghdatis
 Football
 Chelsea vs Barcelona
 Werder Bremen vs Levski Sofia
 Inter vs Spartak Moscow
 Sporting Lisbon vs Bayern Munich
 Bordeaux vs Liverpool
 Galatasaray vs PSV
 Olympiakos vs Roma
 Valencia vs Shakhtar
 Basketball

Figure 1. Screen dump of betting in running on cricket

5. Sportsbet21

Early in 1998 Mark Solonsch of Synaval Pty Ltd approached Swinburne Sports Statistics with an idea of developing an automated betting system to allow real time bets on events that occur during a sporting event. The initial idea was to bet on the number of runs in an over of cricket. The system would be driven by computer-generated odds that would take into account the state of the game and the past betting pattern of the punters.

As discussed in section 2, there are several problems with sports betting. Consider one-day cricket. With approximately 100 international matches per year, betting opportunities are limited. This contrasts with horse and harness racing where there might be that many races in one day in Australia. A bet takes a day (or 5 days in test cricket) to decide, and this reduces churn, the opportunity for punters to 'reinvest' their winnings, reducing turnover for the bookmaker. Furthermore, with head to head betting there are only two outcomes, and this restricts the bookmaker to a margin of about 5%. A small number of bets and a small margin both decreases profit and increases the risk for the bookmaker. On the other hand, betting on the number of runs in an over of cricket provides 100 bet opportunities in one match, and with over 10 outcomes allows higher betting margins than head to head. Difficulties to be overcome included the small betting time between overs to set odds and take and validate bets. This necessitated the development and testing of a suitable statistical model that would set the odds virtually instantly, and suitable computer protocols to allow a large number of bets in a small time interval.

A statistical model was developed by analyzing data from all one day matches played up to that time. Similar to the analysis described in Section 3, relevant variables were determined by a statistical analysis. Logistic regression methods were used to estimate the probability of scoring less than or at least each of 11 outcomes, being the number of runs from 1 to 10 and more than 11 runs. Variables used included scores, run rates and wickets fallen over various periods, over number and target score. Obviously separate models had to be developed for the first and second innings. Bailey & Clarke (2006) gives another example of the sort of statistical modeling undertaken, in this case to predict the chances of each team ultimately winning the game. Once true probabilities were predicted, these were converted to bookmaker's odds, allowing for any given percentage profit. As soon as the operator entered the score for the over, the model virtually instantly calculated the prices for each of the 11 outcomes for the next over. This allowed punters approximately 30 seconds to place bets before betting was closed at the start of the next over.

Live Betting

Cricket Tennis
 T Henman vs J Nieminen

Stake Bet now

Taking bets now!
 Match betting
 Game 9, Henman serves

Selection	Odds	Select
Server wins to love	4/1	<input type="checkbox"/>
Server Win to 15	5/2	<input type="checkbox"/>
Server wins to 30	3/1	<input type="checkbox"/>
Server wins after deuce	7/2	<input type="checkbox"/>
Receiver wins to love	33/1	<input type="checkbox"/>
Receiver wins to 15	100/7	<input type="checkbox"/>
Receiver wins to 30	10/1	<input type="checkbox"/>
Receiver wins after deuce	12/1	<input type="checkbox"/>

Game	score
11	-
10	-
9	-
8	-
7	15 - Win
6	30 - Win
5	Win - Deuce
4	Love - Win
3	Win - 15
2	Love - Win
1	Win - 30

Figure 2. Early screen dump of betting in running on tennis

In early 2000 a trial was run with volunteer punters given a certain amount of play money to bet on a one-day match. The results were positive in that they achieved a return to the operators and positive feedback from players. The startup company Sportsbet 21 was set up with Swinburne and staff holding 50%, and Synaval and staff holding 50%. Subsequently Tasmania Tote bought a share of the company, which provided funds for further development work. The system underwent several further trials, and a License agreement was finally struck with Ladbrokes. The system was run live with Ladbrokes under the name 'betweenovers' during the 2003 World Cup. The product achieved profit targets on reasonable turnover, and demonstrated the robustness of the mathematical models. Figure 1 shows a recent screen dump of the product. The product has continued to operate profitably under the name 'bet in play', although the limited time allowed for betting has been a restricting factor. Subsequent developments have allowed for setting prices an over in advance, so prices on over 21 (say) will be set and betting commence as soon as over 20 starts. Models have also been developed for test cricket and partnership length, and 20/20 cricket is on the drawing boards.

With the success of the cricket model, it was decided in 2003 to extend into tennis by setting prices on each game score. Thus the server could win to love, 15, 30, or deuce, or lose to love, 15, 30, or deuce, giving 8 outcomes and approximately 10 bets in each set of a tennis match. The modeling approach taken differed significantly from that taken in cricket. Firstly, the model was player specific, and the past player statistics of the two participants were taken as input data. Barnett & Clarke (2005) discuss the player statistics available, and how they can be used to calculate other statistics of interest. Secondly, a single parameter (chance of server winning a point) was used as the input to a probability model giving the chances of all possible outcomes. This single parameter was updated as the match progressed. Since the prices for a server's next service game were determined as soon as he finished serving his current game, the system always allowed for betting a game ahead. This has allowed for betting to be continually open. Figure 2 shows the Ladbrokes screen for tennis betting. Recently the ability to bet on the winner of the game has been added to the product.

Acceptance by the public of betting in running continues to grow, with the models generally standing up to punters' acumen.

Conclusion

Benter et al (1996) and Shepherd (2001) showed that modeling could be used to successfully exploit inefficiencies in the traditional horse and dog racing markets. Dixon & Coles (1997) and Goddard & Asimakopoulos (2004) showed similar inefficiencies exist in soccer betting. Given that sports betting punters may be less sophisticated, and that house percentages are often less than a third of those in racing, inefficiencies should be easier to exploit. The examples given here are not just theoretical profits derived from post event analyses, but include examples of actual profits made in real time betting. Quantitative models can be used to counteract the passion that is often used to drive betting on team and individual sports. Methods do not have to be overly

complex, and we have shown that quite simple prediction models can produce profits. With the growing range of sports betting products, opportunities also exist for mathematical models to assist or replace the bookmaker's function of setting odds. □

Acknowledgements

We wish to acknowledge the contributions of colleagues and associates to the work described here. Our thanks to Tristan Barnett, Rod Clarke, Myles Harding, Mark Lowy, and Mark Solonsch.

REFERENCES

- Bailey, M. (2000). Identifying arbitrage opportunities in AFL betting markets through mathematical modelling. In *Proceedings of the Fifth Australian conference on Mathematics and Computers in Sport* (G. Cohen & T. Langtry Eds.). University of Technology Sydney: Sydney. pp. 37-42.
- Bailey, M. (2005) Predicting sporting outcomes: A statistical approach: Unpublished PhD thesis, Swinburne University, Melbourne.
- Bailey, M. & Clarke, S. R. (2002). Predicting the Brownlow medal winner. In *Proceedings of the sixth Australian conference on mathematics and computers in sport* (G. Cohen and T. Langtry Eds.). University of Technology, Sydney: Gold Coast. pp. 56-62.
- Bailey, M. & Clarke, S. R. (2004a). Deriving a profit from Australian rules football: a statistical approach. In *Proceedings of the Seventh Australian conference on Mathematics and Computers in Sport* (H. Morton, Ed.). Massey University: Palmerston Nth. pp. 48-56.
- Bailey, M., Miel, G. & Clarke (2004b). Market Inefficiencies in Player Head to Head Betting on the 2003 Cricket World Cup. In *Economics, Management and Optimisation in Sports* (P. Pardalos, S. Butenko et al, Eds.). Springer. pp. 185-201.
- Bailey, M. & Clarke, S. R. (2006). Predicting the match outcome in one day international cricket matches, while the game is in progress. *Journal of Sports Science and Medicine* 5, 480-487
- Barnett, T. J. & Clarke, S. R. (2005). Combining player statistics to predict a long tennis match at the 2003 Australian Open. *International J. of Management Mathematics*. 16, 113-120.
- Benter, W., Miel, G. & Turnbull, P.D. (1996) Modelling distance preference in thoroughbred racehorses, in *Proceedings of Third Conference on Mathematics and Computers in Sport* (N. de Mestre, Ed.). Bond University: Queensland. p. 207-218.
- 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. & Clarke, R. C. (2006). A Comparison of computer and human predictions of AFL. In *Proceedings of the Eighth Australasian conference on Mathematics and Computers in Sport* (J. Hammond & N de Mestre Eds.). MathSport (ANZIAM): Coolongatta. pp 141-149.
- Dixon, M.J. & Coles, S.G. (1997). Modelling Association football scores and inefficiencies in the UK football betting market. *Applied Statistics*. 46, 265-280.
- Elderton, W. (1927). *Frequency Curves and Correlation*. London., Layton.
- Goddard J. & Asimakopoulos, I. (2004) *Journal of Forecasting*. 23 (1): 51-66
- Haigh, J. (1999). *Taking chances: Winning with Probability*. New York, Oxford University Press.
- Kelly, J. (1956). A New Interpretation of Information Rate. *Bell System Technical Journal*, 35, 917-926.
- Kimber, A. C. & Hansford, A. R. (1993). A statistical analysis of batting in cricket. *Journal of the Royal Statistical Society (Series A)*. 156(3): 443-445.
- Morton, R. H. (2002). Team rating and home advantage in SANZAR rugby union, 2000-2001. In *Proceedings of the Sixth Australian conference on Mathematics and Computers in Sport* (G. Cohen & T. Langtry Eds.). University of Technology Sydney: Sydney. pp. 182-187.
- Shepherd, C. (2001). Discrete choice modelling in a small betting market, Unpublished PhD Thesis, Department of Econometrics and Business Statistics. Monash University: Melbourne.
- Wood, G. H. (1945). Cricket scores and geometrical progression. *Journal of the Royal Statistical Society (Series A)*. 108: 12-22.
- Yelas, S. (2003). Analysis of Swinburne computer performance in identifying inefficiencies in the AFL betting market, Unpublished MSc Thesis, School of Mathematical Sciences, Swinburne University: Melbourne.
- Yelas, S. & Clarke, S. R. (2004). Forecasting the 2003 Rugby World Cup. In *Proceedings of the Seventh Australasian conference on Mathematics and Computers in Sport* (H. Morton Ed.). Massey University: Palmerston Nth. pp. 270-277.