The Statistician (1994)
43, *No.* 2, *pp.* 309–315

Focus on Sport

Index betting on sports

By DAVID A. JACKSON†

Trinity College, Dublin, Republic of Ireland

[Received September 1992. Revised June 1993]

SUMMARY
The rapid growth of sports betting in Europe beginning in the mid-1980s has continued into the 1990s. The gaming industry has been slow to take full advantage of this opportunity owing to a lack of implementation of management science and operations research techniques by managers. Additionally, recruitment of odds compilers with professional qualifications by the major bookmakers in the UK is practically unknown. Index betting on sports is an important new area of sports betting. Examples from American football, soccer and tennis are quoted and the main differences between a sports index and a stock-market index are discussed. All the companies competing for a share of the index betting market in the UK have experienced difficulties due to their lack of management science and operations research expertise. Some of the basic modelling techniques that the managers of these companies should have at their disposal are illustrated by using estimates from Wimbledon to calculate an index for the total games in a tennis match.

Keywords: Gaming; Index betting; Modelling; Odds compiler; Sports

1. Sports betting

Sports betting is the fastest growing area of the gaming industry in the UK. The major bookmakers make a distinction between their core business which is betting on horse- and greyhound racing, and sports betting which includes betting on soccer, tennis, rugby, American football, snooker and even political elections. The largest of these bookmaking companies is Ladbrokes with over 2100 High Street outlets in the UK alone, and extensive gaming or gaming-related interests in many other countries, including the USA. In 1986 I had an interview with the Managing Director of Ladbrokes's racing division, to obtain some funding for research into sports betting. Sports betting at that time, he told me, accounted for only 1% of his company's betting turnover. The figure for 1990 was just over 10%, for 1992 it was nearly 20% and it seems set to rise considerably in future years. A further indicator of a possible boom in sports betting in the UK is provided by a comparison with the situation in the USA. The 1989 figures for Nevada, which is the only state where bookmaking is legal, were sports betting \$1.4 billion (78%), horse- and greyhound racing \$0.4 billion (22%).

The experts who make the prices for Ladbrokes and the other major bookmakers are known as odds compilers. These odds compilers in sports betting are in general early school-leavers who have shown promise in the cash betting offices, and then been given some in-house training. I enquired whether Ladbrokes had ever considered recruiting their odds compilers from graduates. The Managing Director's reply was 'God forbid'. Needless to say that particular request for funding was not a success. However, perhaps the rapid increase

† *Address for correspondence*: Department of Statistics, School of Systems and Data Studies, Trinity College, University of Dublin, Dublin 2, Republic of Ireland.
E-mail: IN% "DJACKSON@VAX1.TCD.ie"

 0039–0526/94/43309

in sports betting which is taking place now, and some of their experiences in the intervening years, will have persuaded Ladbrokes to change their recruitment policy. Hopefully, in addition to employing a few statisticians, they may even contemplate funding a little university research.

There are perhaps as many different ways to bet on sports as there are sports to bet on. A newcomer to the scene is index betting. In this paper I outline the nature of index betting on sports and show, by analysing some basic models, how an index can be calculated. At the same time I hope to make the point that if the gaming industry wishes to take full advantage of the current boom in sports betting it will have to acquire a little more academic expertise.

2. Index betting

Index betting on sports is a new area. However, index betting itself is not new. It has been around for quite some time on the futures markets of stock exchanges around the world, where, in essence, investors gamble on the future price of various commodities. In recent years several companies have been formed, such as City Index, Spread Bet International (SBI) and I. G. Index, who have extended the concept of index betting from the commodity market to the sports arena and who specialize in quoting an index on a wide variety of sports.

2.1. *Some examples of index betting on sports*

2.1.1. *Spread Bet International: soccer World Cup finals, Italy, July 1990, England versus Ireland.*

	Index
Total goals scored in match	2.1–2.4
Margin of superiority, England over Ireland	0.85–1.1
Result	England 1; Ireland 1

Here SBI are treating the total goals that will be scored in the match as a commodity and one can *buy* this commodity from them at 2.4 or *sell* to them at 2.1. Similarly, the margin by which England's score in goals will exceed Ireland's is treated as a commodity for which SBI quote a two-way price. The index of 0.85–1.1 for England over Ireland indicates that England are favoured to beat Ireland and that SBI expect the margin of their victory to be about 1 goal.

2.1.2. *City Index: American football Superbowl XXIV, New Orleans, January 1990, San Francisco versus Denver.*

	Index
Total points scored in game	44–46
Margin of superiority, San Francisco over Denver	11–12
Result	San Francisco 55; Denver 10

2.1.3. *City Index: tennis, Wimbledon final, July 1990, Becker versus Edberg.*

	Index
Total games in match	40–42
Margin of superiority Becker over Edberg	2–3
Result	Edberg beat Becker 6–2, 6–2, 3–6, 3–6, 6–4

i.e. total games, 44; margin Becker over Edberg, −4.

2.2. *Profit and loss*

Typical bets on the Becker–Edberg Wimbledon tennis final might be

(a) SELL *total games* at 40 for £100 per game;
result *total games*, 44;
LOSS = 4 × £100 = £400;
(b) SELL *margin of superiority* Becker over Edberg at 2 for £100 per game;
result *margin of superiority*, −4;
PROFIT = 6 × £100 = £600.

In most forms of betting the amount that a client can lose, his stake, is fixed. The amount that he can win can be fixed or variable. As index betting on sports is relatively new, it is worth emphasizing that exposure to *loss* as well as potential *profit* are both variables depending not just on whether a particular value is exceeded or not, but also by how much it differs from that particular value.

3. Index betting on sports *versus* stock-market

There are many obvious similarities between index betting on sports and gambling on the future prices of commodities on the stock exchange. However, there are also some major differences.

3.1. *Difference 1: in sport the event takes place*

In Fig. 1 I have charted three possible realizations of the same sports index when the event takes place and contrasted these with the usual behaviour of a commodity index. The sport represented is American football and the final values represented are the total points scored in each of the last three Superbowls.

The index on a sporting event may undergo small changes in the period before the event. Injuries, weather conditions and market forces will all have an effect. However, when the event takes place, the outcome determines a final value for the index, often considerably different from the range of values at which it was bought and sold. For a commodity index, although some unpredicted catastrophe may cause a large fluctuation in the index, no such discontinuous behaviour is expected.

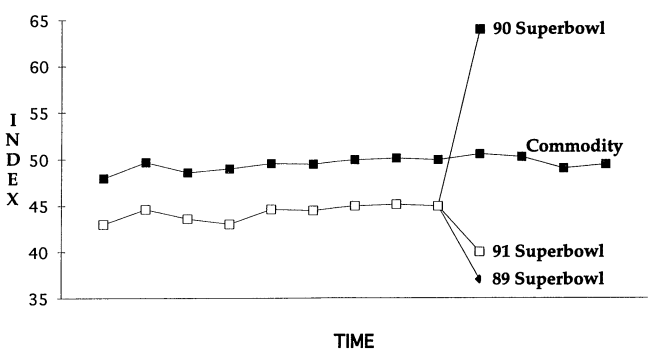

Fig. 1. The final value for a sports index, determined when the event takes place, may differ considerably from the range at which it was traded: this contrasts sharply with the usual behaviour of a commodity index

In theory this difference in behaviour just means a large profit or loss on particular bets. In practice these new companies have exposed themselves and their clients to quite enormous losses and are generally unfamiliar with the credit controls necessary to implement index betting on sport. Lack of knowledge regarding the expectation and, more seriously, the variance for the final value of the index on which they are trading has led to some severe credit losses for several of these index betting companies.

3.2. *Difference 2: modelling*

In sport we can construct good basic models, often depending on only one or two parameters, which describe the underlying process with reasonable accuracy. Sports data are usually plentiful and more importantly reliable, allowing for good estimation and fitting procedures for these theoretical models. This contrasts with the stock-market, where models which describe the underlying process are scarce, tend to be extremely complex and are generally understood by only a small percentage of the financial community.

There are benefits to anyone in the gaming industry, not just those involved in index betting, of an accurate model of the sport on which they are quoting prices. In particular there is the competitive edge that follows from a better understanding of the underlying process. But at present the major bookmakers and the smaller index betting companies suffer from a lack of academic expertise. For some of the index companies this has been disastrous. Although several new companies started up in 1992, SBI ceased trading in August 1991, I. G. Index abandoned sports to concentrate on the financial markets, whereas City Index, originally the market leader in this form of gambling, have drastically reduced the sports betting side of their operation. The remedy is simple.

4. Basic models

4.1. *Poisson models*

Many sports involving two teams, such as soccer or ice-hockey, can be modelled as a Poisson process whereby teams A and B score goals independently at rates λ_A and λ_B respectively. The two parameters describe the process.

For instance the variable total goals is just the *sum* of two Poisson random variables and the variable margin of superiority is just the *difference* of the same two variables:

total goals (expected value) $= \lambda_A + \lambda_B$;
margin of superiority A over B (expected value) $= \lambda_A - \lambda_B$

The variance of both these quantities is $\lambda_A + \lambda_B$.

Direct estimates of the parameters λ_A and λ_B can be made from data concerning each of the two teams. It is not necessary to restrict oneself to matches between the two teams. Morrison and Wheat (1986) used this method for their ice-hockey data. In soccer, rather than to estimate the scoring rates directly, it is more common to use the data to estimate the probabilities for the three possible outcomes to a match, namely

$$P_A = P(\text{A wins}), \qquad P_B = P(\text{B wins}), \qquad P_T = P(\text{tie})$$

where

$$P_A + P_B + P_T = 1.$$

Obviously, for a Poisson process, knowledge of the parameters will uniquely determine the probabilities P_A, P_B and P_T. It is also true that, when the underlying process is Poisson, knowledge of the probabilities uniquely determines the parameters λ_A and λ_B.

For the England *versus* Ireland World Cup match in Italy, July 1990, the estimates of the probabilities (a consensus of expert opinion) for the possible outcomes were

$$P(\text{England wins}) = 0.50, \quad P(\text{Ireland wins}) = 0.21, \quad P(\text{match tied}) = 0.29$$
$$\Leftrightarrow \lambda_{\text{ENG}} = 1.29, \quad \lambda_{\text{IRE}} = 0.73.$$

This leads to indices, for total goals of 2.02 and for margin of superiority for England over Ireland of 0.56, each with a variance of 2.02. Comparing these figures with the indices offered by SBI on this match (example 1), we see that total goals is similar, but their quote of 0.85–1.1 for margin of superiority is essentially inconsistent with the probabilities associated with this match. *Selling* England at 0.85 goals therefore appeared very attractive at the time, especially to an Irishman. Since the result of the match was a 1–1 draw, this led to a profit of 0.85 points.

This type of model can easily be extended to sports, such as American football or rugby, which have more than one type of scoring play. Although, for many sports, more sophisticated versions of these models have already been proposed and analysed, the major bookmakers continue to quote prices without personnel who are capable of building or analysing models at even the most basic level.

4.2. *Binomial-type models*

Many sports use a format whereby a contest between individuals or teams is decided not by a single trial but by a series of trials, supposedly identical. Tennis with a best of three or five sets contest is an obvious example, as is snooker where best of 17 frames matches are commonplace.

If our interest is in S = total sets in a best of five sets tennis match, a simple model is 'P is the probability that a player wins a single set and sets are independent—the single parameter P describes the process':

(a) for $P = 0.5$, $E(S) = 4.125$ and $\text{var}(S) = 0.609$;
(b) for $P = 0.7$, $E(S) = 3.895$ and $\text{var}(S) = 0.623$.

However, if our interest is in total games in a set, then to construct a model we must recognize the fundamental nature of a set of tennis, namely that there are two different types of game in a set, corresponding to when player A is serving and when player B is serving. The model is then 'let P_{A} and P_{B} be the probabilities for players A and B respectively of winning a game when serving: games are independent and service alternates'.

This is a binomial-type model (George, 1973; Riddle, 1988), and we can calculate, purely in terms of the two parameters, the probabilities that player A will win or lose by any of the scores 6–0, 6–1, ..., 6–4, 7–5 or that the score will reach 6–6 in which case a tie-breaker is played.

(Both of these simple models have included the assumption of independent trials and for some sports this assumption may be reasonable. For instance Elo (1978) stated that, for several world championships of chess, in which a series of trials is played, the numbers of wins, losses and draws follow closely the expected distribution under independence. However, for many sports, such as tennis, where the time interval between trials is not of the order of a day or two, as it is in chess, the assumption of independence cannot easily be defended (Jackson, 1993). I have found that models which include effects due to the score, such as psychological momentum, i.e. 'success breeds success', 'failure breeds failure', have proved more accurate.)

For any value of P(probability of winning a set), it is possible to construct a chart showing the expected set length in games, for those points $(P_{\text{A}}, P_{\text{B}})$ which give rise to

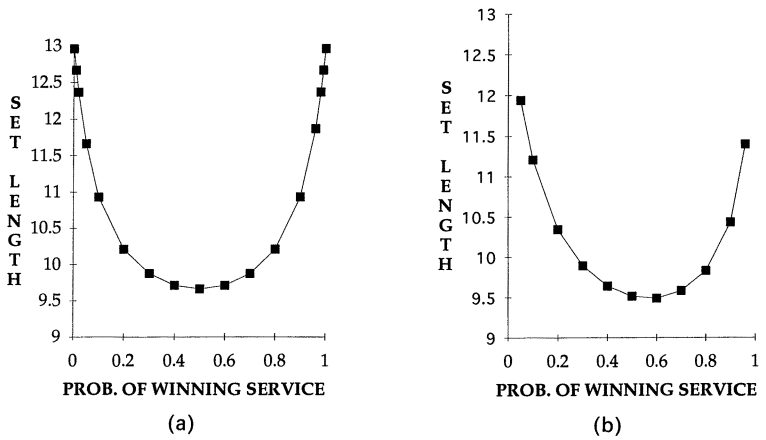

Fig. 2. Expected set length in games plotted against P_A(probability that player A wins his service game) for (a) players of equal strength, i.e. $P = 0.5$, and (b) $P = 0.7$

that value of P. Fig. 2 shows expected set length for $P = 0.5$ (Fig. 2(a)) and $P = 0.7$ (Fig. 2(b)).

For the minimum expected set length:

(a) $P = 0.5$ minimum for

$$P_A = P_B = 0.5 \Rightarrow E(\text{set length}) = 9.66 \text{ and var(set length)} = 3.70;$$

(b) $P = 0.7$ minimum for

$$P_A = 0.57\ (P_B = 0.43) \Rightarrow E(\text{set length}) = 9.49 \text{ and var(set length)} = 3.74.$$

In both these cases, and this is true generally, for any fixed value of P(probability of winning a set), the minimum expected set length in games occurs when the probability for each player of winning a game does not depend on who is serving, i.e. $P_A = 1 - P_B$. As service becomes more important, the expected set length increases. For example, in case (a) where $P = 0.5$ if the situation was that both players had a probability of 0.9 of winning a game when serving, then we can see from Fig. 2 that the expected set length will be over 11 games.

4.3. *Estimates from Wimbledon tournament*

In general, estimates for the parameters P_A and P_B in a particular match will depend on several factors, apart from the level of ability of the two players, such as the playing surface and weather conditions. Estimates for professional players (for whom service is most definitely an advantage) based on data from the Wimbledon tournaments in 1987 and 1988 are as follows:

(a) $P = 0.5$, Wimbledon, $P_A = P_B = 0.8 \Rightarrow E$(set length) = 10.21 games (variance 3.6), $\Rightarrow E$(match length) = 42.1 games (variance 127);
(b) $P = 0.7$, Wimbledon, $P_A = 0.83$, $P_B = 0.72 \Rightarrow E$(set length) = 9.95 games (variance 3.72), $\Rightarrow E$(match length) = 38.5 games (variance 120).

The expected values for match length have been calculated by combining the model for the number of sets in a best of five sets match with the model for the number of games in a set and using the independence properties.

5. Discussion

Some basic models for various sports have been introduced to demonstrate the modelling process. I do not suggest that all these models provide the best fits to the data that they are trying to describe. In practice the simple Poisson models for soccer and American football are reasonable, but for the binomial-type models the assumption of independent trials is often not justified and leads to a poor fit for many sports, tennis in particular. The more sophisticated models, to which I have referred, which allow probabilities to vary according to the score, usually add only one more parameter. For tennis these dramatically improve the fit of the models to available data and lead to lower estimates for, say, total games in a tennis match.

There are major differences between index betting on sports and gambling on the future prices of commodities. Firstly a large variance for the final value of a sports index is common and some knowledge of its magnitude is essential if credit control problems are to be avoided for both the company and the client. Secondly simple and fairly accurate models are available for many sports which can eliminate some of the inconsistencies that exist at present. Hopefully it has been demonstrated that some investment in qualified personnel capable of modelling sports will pay dividends for those companies who wish to take advantage of what is clearly a boom area in the gaming industry.

References

Elo, A. E. (1978) *The Rating of Chessplayers, Past and Present*. London: Batsford.
George, S. L. (1973) Optimal strategy in tennis: a simple probabilistic model. *Appl. Statist.*, **22**, 97–104.
Jackson, D. A. (1993) Independent trials are a model for disaster. *Appl. Statist.*, **42**, 211–220.
Morrison, D. G. and Wheat, R. D. (1986) Pulling the goalie revisited. *Interfaces*, **16**, no. 6, 28–34.
Riddle, L. H. (1988) Probability models for tennis scoring systems. *Appl. Statist.*, **37**, 63–75.