

Winning the Coin Toss and the Home Team Advantage in One-Day International Cricket Matches

Basil M. de Silva and Tim B. Swartz *

Abstract

This paper provides a statistical analysis of 427 one-day international cricket matches played during the 1990's. Two general conclusions are obtained: (1) contrary to widespread opinion, winning the coin toss at the outset of a match provides no competitive advantage and (2) the advantage of playing on one's home field increases the log-odds of the probability of winning by approximately .5.

Keywords : ODI cricket matches, statistical modelling.

*B de Silva is Senior Lecturer, Department of Statistics and Operations Research, RMIT University, GPO Box 2476V, Melbourne Victoria, Australia 3001. T Swartz is Associate Professor, Department of Mathematics and Statistics, Simon Fraser University, Burnaby BC, Canada V5A1S6. T. Swartz was partially supported by a grant from the Natural Sciences and Engineering Research Council of Canada. The authors wish to thank all 3 referees for insightful comments that led to an improved paper.

1. INTRODUCTION

The game of cricket has world-wide appeal and is enjoyed by people of all ages. Despite this fact and despite the intense bookkeeping and scrutiny that the game has received, there has not been an overwhelming amount written on the sport from a statistical perspective. Hopefully, a close to complete list of papers includes Crowe and Middledorp (1996), Kumar (1996), Ganesalingam, Kumar and Ganeshanandam (1994), Kimber (1993), Clarke (1991), Danaher (1989), Chapman (1987), Burrows and Talbot (1985), Croucher (1982), Sumner and Mobley (1981), Pollard (1977), Pollard, Benjamin and Reep (1977), Elderton (1945) and Wood (1945).

In this paper we consider a statistical analysis of two issues associated with one-day cricket matches. In first-class cricket (3 or more days), winning the coin toss is undoubtedly regarded as a great advantage. Since the one-day match is the modern offspring of the first-class game, it might casually be assumed that winning the coin toss in the one-day game is also an advantage. We investigate this belief. We also investigate the effect of the home-team advantage in one-day matches. Both of these topics are controversial and we believe that data can shed some light on their resolution.

Our data consists of one-day international (ODI) matches involving games between the 9 nations belonging to the International Cricket Council (ICC). These games represent the game of cricket played at the highest level with relative stability amongst the teams. We have collected data on the 427 matches played during the 1990's up until the Asia Cup concluding in July 1997. This time period captures the modern game of cricket where the rules have been relatively uniform. It is also the case that recent data is more extensive and reliable. To keep strategies constant, we have limited the data to full 50-over matches and have ignored matches decided by run rates. The data was collected from the comprehensive CricInfo web page (<http://www.cricket.org>).

In Section 2, we look at data involving a team's choice to bat first upon winning the coin toss at the outset of a match. We come to the surprising conclusion that winning the coin toss

provides no competitive advantage. In Section 3, we consider the question of the existence of a home team advantage. We conclude that such an advantage does exist and we quantify the advantage. In Section 4, we provide some concluding remarks.

2. BATTING ORDER STRATEGIES

At the beginning of a match, a coin is tossed and the team that wins the toss is granted the choice of batting first or second. Some people believe that a team should bat first, establish a number of runs and produce a psychological hurdle for the second team to overcome. Other people believe that there is an advantage in batting second as this team knows what score its opponent has produced. This additional information allows the team batting second to adjust their strategy accordingly. Still others feel that the choice between batting first or second should depend on auxilliary and subjective variables such as the weather, the pitch (ie. field) conditions, the team's health, the team's morale, the opponent, whether the team will bat in daylight or under floodlights, etc. Clearly, this is a topic of considerable interest.

As a preliminary study, Table 1 provides summary data on the 427 ODI matches involving the 9 ICC nations. We see from column B_1 that there is great disparity amongst the various teams with respect to their decision to bat either first or second. For example, upon winning the coin toss, Australia chooses to bat first 87% of the time whereas Sri Lanka chooses to bat first only 36% of the time.

Ideally, we would like to compare various *proposed* strategies. However, in most cases, such strategies would depend on variables (eg. team morale, coaching whims, pitch conditions) that can not be easily extracted from past data. Another difficulty in comparing proposed strategies is that matches have not been randomized according to these strategies. Therefore perceived effects due to proposed strategies may actually be attributed to an unknown confounding variable. For example, consider the very simple strategy that New Zealand should always bat first. Even though New Zealand has chosen to bat first in roughly half of its matches, it may be the case that they did so only in matches that they thought they should win. In this case, batting first

would be confounded with games involving weaker opposition.

For these reasons, we only investigate *observed* strategies of the 9 ICC teams. In this context, we would like to compare a team's strategy against its opposite strategy. However, this too, can not be tested. The reason is that data on the opposite strategy is unavailable since we do not know what strategy is being incorporated when a team loses the coin toss. Therefore, we are left with comparing a team's strategy versus its opponents' opposite strategies. More precisely, we look at a team's performance when it wins the coin toss versus its performance when it loses the coin toss.

Consider the data (x_i, y_i) , $i = 1, \dots, n$ where n is the number of games played by the team of interest, $x_i = 1(0)$ if the team wins(loses) the coin toss in the i^{th} game and $y_i = 1(0)$ if the team wins(loses) the game. We have the statistical model $y_i | x_i = 1 \sim \text{Bernoulli}(p_i)$ and $y_i | x_i = 0 \sim \text{Bernoulli}(q_i)$ where $P(x_i = 1) = P(x_i = 0) = 1/2$ for $i = 1, \dots, n$. We are therefore interested in comparing the strategy $\sum_{i=1}^n p_i$ versus the strategy $\sum_{i=1}^n q_i$. We note that if the opponents of a team collectively have the team's opposite strategy, then $\sum_{i=1}^n p_i = \sum_{i=1}^n q_i$. Of the 427 matches in the data set, 8 games resulted in ties. We exclude these matches from the analysis.

The difficulty with the above model is that we have $2n$ parameters (ie. p_i 's, q_i 's) and only n observed responses (ie. y_i 's). We therefore consider several methods to compare $\sum_{i=1}^n p_i$ with $\sum_{i=1}^n q_i$. The first method requires the assumption that $p_i = p$ and $q_i = q$ for all $i = 1, \dots, n$. This simplistic assumption is clearly unrealistic as it assumes that the opponents are all of equal strength and that the playing conditions are constant over time. However, it is a good starting point and the test of $H_0 : p \leq q$ versus $H_1 : p > q$ is easily carried out using a two-sample Binomial test. The first binomial variable is the number of wins having won the coin toss and the second binomial variable is the number of wins having lost the coin toss. The p-values for each of the ICC nations are given in column 1 of Table 2. These p-values are very high; to be significant, p-values must be small (typically $< .05$). Therefore, under this model, we observe no evidence of successful strategies for any of the 9 ICC teams.

Our second method allows for opponents of varying strengths. However, this approach is also unrealistic since it assumes that teams have constant strength over time. We model $\text{logit}(p_i) = \alpha + \beta x_i + \gamma_i$ where $\gamma_i = \gamma_k$ if the i th game is against opponent k , $k = 1, \dots, 8$ and $\sum_{k=1}^8 \gamma_k = 0$. The p-values for testing $H_0 : \sum_{i=1}^n p_i \leq \sum_{i=1}^n q_i$ versus $H_1 : \sum_{i=1}^n p_i > \sum_{i=1}^n q_i$ for each of the ICC nations are based on logistic regression and are given in column 2 of Table 2. The p-values were obtained using the function `glm` (binomial family) in S-PLUS (Venables and Ripley (1994)). Based on tests at level .05 significance, none of the teams exhibit effective strategies. The team which comes closest to significance is India (p-value = .13). We note that India chooses to bat first in approximately half of its matches.

Our third method is based on a Bayesian approach. Allowing flexibility for the strengths of the different teams and allowing for varying strengths over time, we express the likelihood as

$$[y, x \mid p, q] \propto \prod_{i=1}^n p_i^{y_i x_i} (1 - p_i)^{(1 - y_i) x_i} q_i^{y_i (1 - x_i)} (1 - q_i)^{(1 - y_i)(1 - x_i)}$$

Introducing independent priors $p_i \sim \text{Beta}(a_i, b_i)$ and $q_i \sim \text{Beta}(c_i, d_i)$, the Bayesian paradigm immediately gives independent marginal posterior distributions $[p_i \mid x, y] \sim \text{Beta}(a_i + y_i x_i, b_i + (1 - y_i) x_i)$ and $[q_i \mid x, y] \sim \text{Beta}(c_i + y_i (1 - x_i), d_i + (1 - y_i)(1 - x_i))$. The posterior probability of H_0 is then $P(H_0 \mid x, y) = P(\sum_{i=1}^n p_i \leq \sum_{i=1}^n q_i \mid x, y)$ which is obtained by simulating from the marginal posterior distributions and calculating the proportion of time that $\sum_{i=1}^n p_i \leq \sum_{i=1}^n q_i$. The tradeoff for the flexibility of the Bayesian model is the need to specify prior information. We consider $\text{Beta}(1, 1)$ priors (ie. $\text{Uniform}(0, 1)$ priors) which are standard default priors and are often used to represent ignorance. Here the posterior distributions of $\sum_{i=1}^n p_i$ and $\sum_{i=1}^n q_i$ reduce to the sums of independent uniform and triangular distributions. Column 3 of Table 2 gives the posterior probabilities for each of the ICC teams. Again we see no evidence of successful strategies.

As a fourth method, it was suggested by a referee that we try an empirical Bayes approach where the prior parameters a_i, b_i, c_i, d_i , $i = 1, \dots, n$ are determined by the data. Using the

observed winning proportions W_o from Table 1, we set $W_o = a_i/(a_i + b_i)$, $i = 1, \dots, n$ where $a_i/(a_i + b_i)$ is the mean prior probability of winning the i^{th} match given a winning coin toss. Similarly, we set $W_o = c_i/(c_i + d_i)$, $i = 1, \dots, n$. It is not so obvious how one might determine the second parameter in the prior Beta distributions. We decided on setting $[\min(1 - W_o, W_o)/3]^2$ equal to the prior variances. We reasoned that ± 3 standard deviations would bring us to the closest endpoint (ie. either 0 or 1) and would therefore provide prior distributions that ought to capture the true parameter values. Column 4 of Table 2 gives the posterior probabilities based on this empirical Bayes approach. Here, there is even less evidence that any of the observed strategies are successful. The posterior probabilities using methods 3 and 4 were obtained using a Fortran program. The IMSL subroutine `rbet` was used to generate the Beta variates.

In summary, under all 4 methods of analysis, we have observed that the team strategies offer no advantage in deciding the outcome of matches. Although the 9 sets of strategies do not include all possible strategies, we suggest that it would be unlikely for someone to discover a better strategy. We therefore conclude that winning the coin toss has no impact on the outcome of matches.

3. THE HOME TEAM ADVANTAGE

The existence of a home team advantage in ODI cricket matches is generally accepted. We provide some statistical evidence for this point of view, and then investigate whether the home team advantage can be quantified.

In many physical sports (eg. American football, soccer, basketball) it is well accepted that a home team advantage exists (see Stern (1997) and Pollard (1986)). Perhaps in these physical sports, the support of the home crowd is such that a physical lift is often provided which carries the home team to victory. For example, in the 1996/1997 season of the National Basketball Association, the home teams won 683 games while losing 506 games for a home team winning percentage of 57.4%.

In sports that are not as physical, a home team advantage may still exist due to familiarity

with local conditions, biased officiating, etc. The game of cricket may fall into this class. It is interesting to note from columns W_h and W_o of Table 1 that every ICC nation has a higher winning percentage during home games. Using the sign test, this is convincing evidence of the existence of a home team advantage (ie. p-value = $1/2^9 \approx .002$).

To investigate the effect of the home team advantage, we modify our notation and let p_{ijk} be the probability that team i defeats team j at site k where $i, j, k = 1, \dots, 9$, and, in addition, $k = 0$ denotes a non-ICC site. We introduce the model $\text{logit}(p_{ijk}) = \tau_i - \tau_j + \gamma_{ijk}$ where $\sum_{i=1}^9 \tau_i = 0$ and

$$\gamma_{ijk} = \begin{cases} \gamma & \text{if team } i \text{ is the home team} \\ 0 & \text{if the game is played on a neutral site} \\ -\gamma & \text{if team } j \text{ is the home team} \end{cases}$$

This is a 9-parameter model where τ_i is a measure of the differential strength of team i . Therefore, the offset $\tau_i - \tau_j$ represents the advantage in log-odds that team i has over team j . The model also assumes that the home team advantage γ is constant over all ICC teams. Note that the logit transformation of p_{ijk} is natural in two respects. Firstly, for teams of equal strength that play on a neutral site, we have $\text{logit}(p_{ijk}) = 0$ which implies $p_{ijk} = .5$. Therefore, there is no need for an intercept term. Secondly, it is sensible to quantify the home team advantage on the log-odds scale since we should expect small relative improvements for strong teams that win most of their games. Conversely, we should expect large relative improvements for weak teams that lose most of their games.

Again we use the function `glm` (family binomial) in S-PLUS to carry out the logistic regression, and again we exclude the 8 tied games from the 427 matches. We obtain $\hat{\gamma} = .53$ with standard error .14. To put this quantity in perspective, a team with a winning percentage of 50% would increase its winning percentage to 63% when playing at home. Therefore the home team advantage provides a considerable edge to the home team.

We now consider a further modification to the modelling procedure. It is generally believed that world-wide, Pakistan and the West Indies are popular teams. This may be due to their flamboyant style (West Indies), their frequency of travel and their high success rates during the late 1980's and early 1990's. For this reason, we consider these 2 nations to have a home field advantage (ie. γ) for matches actually played on neutral sites. Similarly, we consider Sri Lanka and New Zealand to be less popular teams. This may be due to the fact that these nations have small emmigrant populations worldwide. We therefore assign a home team disadvantage (ie. $-\gamma$) to Sri Lanka and New Zealand for matches played on neutral sites. With these modifications, we obtain a comparable $\hat{\gamma} = .50$ with a slightly smaller standard error .12.

4. CONCLUDING REMARKS

Given the increasing world-wide popularity of sport and the increasing access to sports related data, it seems that serious statistical investigations of issues in sport are becoming more common. In 1992, the American Statistical Association formed the Section on Statistics in Sports and have sponsored both publications and sessions. In the same year, Bond University, Australia began biennial meetings on Mathematics and Computers in Sport with Statistics in Sport as one of the stated themes. We also note that since 1988, the statistics magazine *Chance* has devoted considerable attention to sports related topics.

In this paper, we have added to the growing interest in sports and statistics by considering 2 questions involving ODI cricket matches. Surprisingly, we have found that winning the coin toss at the outset of a match provides no competitive advantage. We have also estimated that the log-odds of winning a match increases by approximately .5 when a team plays at home. This second result may have implications for betting.

REFERENCES

- Burrows, B.L. and Talbot, R.F. (1985), "Boycott, Botham and the exponential distribution", *Teaching Statistics*, 7, 42-48.

- Chapman, B. (1987), "An analysis of the determinants of English-Australian test cricket crowds: estimating the value of Don Bradman", Technical Report, Centre for Economic Policy Research, Australian National University.
- Clarke, S.R. (1991), "Consistency in sport - with particular reference to cricket", *NZOR Proceedings*, 27th Annual Conference, Victoria University of Wellington, 30-35.
- Croucher, J.S. (1982), "Anglo-Australian test cricket dismissals 1946-1980", *Bulletin of Applied Statistics*, 9, 179-193.
- Crowe, S.M. and Middeldorp, J. (1996), "A comparison of leg before wicket rates between Australians and their visiting teams for test cricket series played in Australia, 1977-94", *The Statistician*, 45, 255-262.
- Danaher, P.J. (1989), "Estimating a cricketer's batting average using the product limit estimator", *New Zealand Statistician*, 24(1), 2-5.
- Elderton, W.P. (1945) "Cricket scores and some skew correlation distribution", *Journal of the Royal Statistical Society Series A*, 108, 1-11.
- Ganesalingam, S., Kumar, K. and Ganeshanandam, S. (1994), "A statistical look at cricket data", in *Mathematics and Computers in Sport*, 1994, Bond University, Australia, 89-104.
- Kimber, A. (1993), "A graphical display for comparing bowlers in cricket", *Teaching Statistics*, 15, 84-86.
- Kumar, K. (1996), "Is cricket really by chance?", in *Mathematics and Computers in Sport*, 1996, Bond University, Australia, 69-76.
- Pollard, R. (1977), "Cricket and statistics", in *Optimal Strategies in Sports*, S.P. Ladany and R.E. Machol, editors, North Holland: Amsterdam, 129-130.

- Pollard, R. (1986), "Home advantage in soccer: a retrospective analysis", *Journal of Sports Science*, 4, 237-248.
- Pollard, R., Benjamin, B. and Reep, C. (1997), "Sport and the negative binomial distribution", in *Optimal Strategies in Sports*, S.P. Ladany and R.E. Machol, editors, North Holland: Amsterdam, 188-195.
- Sumner, J. and Mobley, M. (1981), "Are cricket umpires biased?", *New Scientist*, 91, no. 1260, 29-31.
- Stern, H.S. (1997), "A statistician reads the sports pages", *Chance*, 10(4), 19-23.
- Venables, W.N. and Ripley, B.D. (1994), *Modern Applied Statistics with Splus*, Springer-Verlag.
- Wood, G.H. (1945), "Cricket scores and geometrical progression", *Journal of the Royal Statistical Society Series A*, 108, 12-22.

Table 1: Summary data for the 9 ICC nations. Here B_1 is the proportion of time that a team chooses to bat first upon winning the coin toss, W_o is the overall winning proportion, W_c is the winning proportion in games that a team has won the coin toss and W_h is the winning proportion in games played on a home field. The quantities in parentheses are the number of cases.

Nation	B_1	W_o	W_c	W_h
Australia	.87 (68)	.63 (127)	.63 (67)	.67 (69)
England	.65 (20)	.36 (45)	.35 (20)	.75 (4)
India	.56 (59)	.49 (106)	.53 (59)	.72 (39)
New Zealand	.55 (51)	.37 (99)	.28 (50)	.50 (44)
Pakistan	.47 (66)	.57 (131)	.56 (64)	.60 (20)
South Africa	.69 (42)	.61 (94)	.55 (42)	.71 (42)
Sri Lanka	.36 (56)	.46 (100)	.46 (56)	.71 (21)
West Indies	.36 (39)	.53 (90)	.58 (38)	.57 (21)
Zimbabwe	.58 (26)	.17 (46)	.17 (23)	.36 (11)

Table 2: P-values and posterior probabilities used to assess the efficacy of strategies involving the decision to bat first. Method 1 requires that the probability of winning a match is constant over opponents and time, method 2 allows for the varying strengths of opponents but requires constant strength over time, method 3 is based on a Bayesian model with independent uniform priors and method 4 is based on an empirical Bayes approach.

Nation	Method ₁	Method ₂	Method ₃	Method ₄
Australia	.53	.38	.48	.50
England	.53	.60	.47	.50
India	.21	.13	.36	.43
New Zealand	.97	.91	.80	.64
Pakistan	.59	.51	.55	.52
South Africa	.85	.91	.71	.58
Sri Lanka	.46	.32	.50	.49
West Indies	.23	.20	.39	.43
Zimbabwe	.50	.40	.50	.50