J. R. Statist. Soc. A (2004)
167, *Part* 4, *pp.* 657–667

Rating teams and analysing outcomes in one-day and test cricket

P. E. Allsopp and Stephen R. Clarke

Swinburne University of Technology, Melbourne, Australia

[Received August 2002. Revised August 2003]

Summary. Multiple linear regression techniques are applied to determine the relative batting and bowling strengths and a common home advantage for teams playing both innings of international one-day cricket and the first innings of a test-match. It is established that in both forms of the game Australia and South Africa were rated substantially above the other teams. It is also shown that home teams generally enjoyed a significant advantage. Using the relative batting and bowling strengths of teams, together with parameters that are associated with common home advantage, winning the toss and the establishment of a first-innings lead, multinomial logistic regression techniques are applied to explore further how these factors critically affect outcomes of test-matches. It is established that in test cricket a team's first-innings batting and bowling strength, first-innings lead, batting order and home advantage are strong predictors of a winning match outcome. Contrary to popular opinion, it is found that the team batting second in a test enjoys a significant advantage. Notably, the relative superiority of teams during the fourth innings of a test-match, but not the third innings, is a strong predictor of a winning outcome. There is no evidence to suggest that teams generally gained a winning advantage as a result of winning the toss.

Keywords: Cricket; Linear modelling; Logistic regression; Sports

1. Introduction

Cricket has two distinct phases: the batting phase and the bowling phase, which provide an indication of a team's attacking and defensive abilities. Generally, the batting team strives to maximize its score by making as many runs as possible whereas the bowling team endeavours to restrict the score of the batting team. A match result can be construed as the combined effect of the batting and bowling abilities of the respective teams. In this paper we shall model the batting and bowling ratings for the International Cricket Council (ICC) one-day and test playing nations as well as gauge which factors associated with a team's first-innings performance in test cricket most critically affect outcomes of matches.

The analysis of the attacking and defensive abilities of sporting teams has been conducted in some sports such as rugby league and soccer but not cricket. We find this surprising given that the batting and bowling phases of cricket are clearly distinct. Lee (1999) used a bivariate negative binomial regression model to account for the attacking and defensive abilities of teams in rugby league and concluded that a team's ability to defend was more important than their offensive capabilities. Dixon and Coles (1997) also incorporated attacking and defensive parameters in a bivariate Poisson model when modelling scores in English soccer and found that the teams that are higher in the league table had the higher average attack and defence ratings.

Address for correspondence: Paul Allsopp, 4 Gledhill Street, Seddon, Victoria, Australia.
E-mail: paulalls@optusnet.com.au

0964–1998/04/167657

One-day cricket is restricted to one innings per team, with each innings allocated a maximum of 50 overs. For completed matches the only outcomes are a win (or loss) or a tie. The team with the higher score is declared the winner. A tie results if both teams attain the same score after completing their innings. A drawn result only arises if weather or other interruptions do not allow a match to be completed to a stage where rain interruption rules are applied.

Test cricket is played between 10 of the ICC cricket playing nations and involves a maximum of two innings per team. A test-match can last up to 5 days. The possible outcomes are a win (or loss), a draw and very rarely a tie. A team wins by having a higher aggregate score after dismissing their opponents in the second innings. If both teams complete their second innings with the same aggregate score the match is a tie. Otherwise the match ends in a draw.

We have considered five seasons of results from 1997 to 2001, which include 248 one-day matches and 151 test-matches. The data were obtained via the Cricinfo archives at `www.cricinfo.com`. Bangladesh has not been included in the test cricket analysis because it has only recently been adopted as an ICC-sanctioned test playing nation and consequently has played in relatively few test-matches. Bangladesh, Kenya and Scotland have not been included in the one-day analysis because they also have played in relatively few matches. Note also that over this period a few matches were played on neutral grounds but these have been excluded from the study. Subsequently, in all matches there is a designated home team.

2. Modelling batting and bowling ratings in one-day and test cricket

In developing batting and bowling ratings for teams in one-day cricket we are faced with an inequity problem when the team batting second wins. The team batting second, unlike the first-innings team, does not expend its resource quota of 50 overs and 10 wickets when it wins a match. Consequently, their scores are truncated and are not a true measure of how well they have performed. To overcome this problem we apply techniques developed by Duckworth and Lewis (1998) and adapted by Allsopp and Clarke (2000) to estimate the projected score of the team batting second after it has exhausted its available resources. The Duckworth and Lewis method sets a revised target for the team batting second when overs in either innings have been lost because of a break in play. The target is revised in accordance with the available run scoring resources in the form of wickets and overs that the two teams have at their disposal at the time of the interruption. By adapting the Duckworth and Lewis method when the team batting second wins a match, we are in effect scaling up its score to allow for the number of further runs that it could make with its unused resources. Duckworth and Lewis have prepared detailed tables from which the resources that are available in the form of wickets and overs are expressed as a single resource percentage R. In addition we shall incorporate modifications that were proposed by de Silva *et al.* (2001) and use modified R-values expressed as $R_{\text{mod}} = (1.183 - 0.006R)R$. This adjustment accounts for the overestimation of predicted 50-over scores when the team batting second uses up very few resources and the underestimation of predicted scores when it uses nearly all its resources. If S_{actual} represents the truncated score at the point of victory, the subsequent projected score for the team batting second is estimated to be

$$S_{\text{projected}} = 100 S_{\text{actual}} / (100 - R_{\text{mod}}).$$

In one-day cricket, if both teams played until all their resources were exhausted it would be expected that the distribution of the scores would be similar. Consequently, the distribution of the second-innings scores after the truncated scores have been replaced with the projected scores should also be similar to that of the first innings. For the first innings, the sample mean and standard deviation are 236.1 and 52.0 runs respectively. For the second innings, with the inclusion of the projected scores, the respective sample mean and standard deviation are 239.4 and 57.4 runs. A test for equal variances confirms that there is no significant evidence to suggest that the variances for the two innings are different ($F = 0.819$; p-value 0.116). Similarly, a two-sample t-test, with assumed equal variances, verifies that the two innings are not statistically different ($t = -0.69$; p-value 0.492). Consequently, there is no valid reason to reject the concept that the estimated second-innings scores are different from what would be expected if second-innings teams could continue batting until their available resources were exhausted.

To rate teams in test cricket we shall focus our attention on the first innings only, where we assume that teams attempt to optimize their performance to establish a substantial first-innings lead. This offers a reliable reflection of a team's batting and bowling abilities since a team's major focus is the establishment of a first-innings lead. Second-innings performances are not being considered because teams tend to be more measured in their approach and accordingly adapt their style of play as a strategic response to what has occurred in the first innings. Consequently, a team's second-innings batting and bowling performances are likely to lack the consistent approach that is more evident in a team's first innings.

A team's first-innings score in a test or an innings score in a one-day match that is played between the batting team i and the bowling team j on ground k is modelled as

$$s_{ijk} = A + a_i - b_j + h_{ik} + \varepsilon_{ijk} \tag{1}$$

where the indices $i, j, k = 1, \ldots, 9$ represent the nine ICC cricket playing nations. Clarke and Allsopp (2001) used similar modelling techniques to analyse the 1999 World Cup of cricket. However, whereas the response variable for the World Cup of cricket data was the margin of victory (or the final differential in runs between competing teams), for the analysis of a team's batting and bowling abilities the actual innings scores will be modelled. The response variable s_{ijk} signifies a team's score, the intercept A represents the expected score between average teams on a neutral ground and a_i and b_j signify the batting and bowling ratings of teams i and j respectively. When $k = i$, the home advantage parameter h_{ik} can be modelled as either a team's individual home advantage h_i or as a common home advantage h. More complicated home advantage models can also be considered. For example, since India, Pakistan and Sri Lanka play in very similar conditions it could be assumed that when these teams played each other neither team enjoyed a home advantage. ε_{ijk} is a zero-mean random error with constant variance. The error term is included because two competing teams will not necessarily repeat their first-innings performances the next time that they meet. Using a design matrix of indicator variables, a least squares regression model is fitted to the scores. For convenience, $\Sigma_{i=1}^{9} a_i = 900$ and $\Sigma_{j=1}^{9} b_j = 900$, which ensure that the average batting and bowling ratings are each 100. Accordingly, ratings above and below this figure provide evidence on how well a team has performed relative to the average performance.

3. Exploratory analysis of test cricket

Tables 1–4 provide a summary of the test-match results over the period of the study. Table 1 provides the overall percentage of wins, draws and losses. Australia and the West Indies have a much lower percentage of draws than the other teams, which may indicate a less defensive style of play. By contrast, New Zealand has a high percentage of draws. Table 2 provides the percentage of wins, draws and losses after a first-innings lead had been established. The consistently strong showing of Australia and South Africa has been well documented, which is clearly evident in the high proportion of matches in which they gain a first-innings lead. All teams except Zimbabwe showed a strong tendency to win after establishing a first-innings lead.

Table 1. Summary of test cricket results for the period 1997–2001

Team	*Matches played*	*% wins*	*% draws*	*% losses*
Australia	42	69	14	17
South Africa	38	55	32	13
Sri Lanka	26	42	23	35
England	44	32	25	43
India	27	30	30	41
West Indies	40	28	15	58
Pakistan	30	27	27	47
New Zealand	30	27	37	37
Zimbabwe	25	12	32	56

Table 2. Summary of test cricket results for the period 1997–2001 after a first-innings lead had been established

Team	*Matches where leading*	*% of matches leading*	*% wins after leading*	*% draws after leading*	*% losses after leading*
Australia	38	90	76	5	19
South Africa	27	71	63	30	7
Sri Lanka	15	58	67	13	20
India	14	52	43	43	14
New Zealand	11	37	64	27	9
West Indies	13	33	62	15	23
Pakistan	13	33	54	23	23
England	14	32	43	50	7
Zimbabwe	6	24	17	33	50

However, only the top four ranked teams established a lead more than 50% of the time. The importance of a first-innings lead is demonstrated by South Africa and New Zealand, which have virtually identical records given a first-innings lead. However, South Africa established a lead almost twice as often as New Zealand. Australia's extremely low percentage of draws after leading contrasts with South Africa's and could indicate a propensity to go for wins even if it risks losing. A fast scoring rate in the first innings would also allow time for both teams to force a win. Of the other teams, Sri Lanka and the West Indies have a lower percentage of draws than losses after leading. By contrast India, New Zealand and England have a high proportion of draws compared with losses after a first-innings lead.

Table 3 allows a comparison of home and away performance in test cricket depending on batting order. The home team won 46% of matches and lost 28%, demonstrating a clear home advantage. However, the home side performed better when batting second. In fact the advantage in batting second was as strong as that enjoyed by the home side, and the home side lost more matches than it won when batting first. The team batting second won $46+28=74$ or 49% of matches and lost only 26%. This is in stark contrast with the accepted wisdom of batting first if the captain wins the toss.

Table 4 provides a summary of the teams' first-innings batting and bowling results. The differentials in the last column provide a measure of the team's average first-innings lead. Table 4 demonstrates the dominance of Australia and South Africa, and the weakness of Zimbabwe, with nearly 100 runs separating them from the rest of the teams.

Table 3. Comparison of home and away performance and batting order in test cricket for the period 1997–2001

	Home team wins	*Home team draws*	*Home team loses*	*Total*
Home team bats first	24	23	28	75
Home team bats second	46	15	15	76
Total	70	38	43	151

Table 4. Batting and bowling result summary for the first-innings of a test-match for the period 1997–2001

Team	*Batting*		*Bowling*		*Differential (batting – bowling)*
	Mean	*Standard deviation*	*Mean*	*Standard deviation*	
Australia	382	116	242	100	140
South Africa	374	116	259	119	115
Sri Lanka	316	125	297	122	19
New Zealand	319	120	329	158	−10
Pakistan	284	96	307	132	−23
India	297	180	333	91	−36
West Indies	237	84	294	108	−57
England	275	116	339	137	−64
Zimbabwe	234	107	391	143	−157

Although summary statistic tables such as these give an indication of relative team strengths, they also show effects such as home advantage, which in turn compromise their usefulness. The test calendar is not balanced—some teams play more series against stronger opponents, and in every series one team has a home advantage. To measure team strengths properly these effects need to be taken into account. We do this in the next section by modelling these effects in one-day cricket and the first innings of test cricket.

4. One-day cricket and test cricket first-innings results

In fitting model (1) to the one-day scores, which includes the replacement of the truncated second-innings scores with the projected scores (as outlined in Section 2), the Anderson–Darling test for normality establishes that the normality assumption is not breached (p-value 0.177). The resultant parameter estimates for the batting and bowling ratings are provided in Table 5 and provide an easily interpretable measure of a team's relative strength in terms of runs. For example, Australia's batting rating of 116 suggests that it performs 16 runs better than expected against an average team, whereas Zimbabwe's bowling rating at 77 suggests that it generally performed 23 runs worse than expected when bowling, so their opponents score 23 runs more than average. Thus Australia batting against Zimbabwe would score $16+23=39$ runs more than average. Similarly Zimbabwe batting against Australia's bowling would score $13+13=26$ fewer runs than average. Hence we would expect Australia to defeat Zimbabwe on a neutral ground by $39+26=65$ runs. This can also be obtained by subtracting the combined

Table 5. Team ratings for one-day cricket for the period 1997–2001

Team	*Batting rating*	*Bowling rating*	*Combined rating*
Australia	116	113	129
South Africa	106	119	125
Sri Lanka	111	106	117
Pakistan	100	101	101
India	114	86	100
England	82	112	94
New Zealand	97	88	85
West Indies	87	98	85
Zimbabwe	87	77	64

rating of Zimbabwe from that of Australia. The combined rating that is given in Table 5 is the sum of the batting and bowling rating, again adjusted to sum to average 100, and gives a measure of a team's overall strength.

The estimate for the common home advantage parameter is 14 runs (p-value 0.002), which suggests that the home team generally gained a significant advantage (in runs). This is consistent with de Silva and Swartz (1997), who showed that a significant common home advantage was evident in one-day cricket. Researchers such as Stefani and Clarke (1992), Clarke and Norman (1995) and Harville and Smith (1994) also considered an individual home advantage effect when modelling margins of victory. This seems reasonable in cricket, where the home advantage that is enjoyed by India might be expected to be different from that of (say) England. However, allowing for individual home advantages, by the addition of eight more parameters, does not provide a significant improvement over the common home advantage model. Various other models for home advantage were tried. For example, since India, Pakistan and Sri Lanka play under similar conditions, we might assume that there is no home advantage when these teams play each other. However, these models also failed to show any significant improvement on the common home advantage model.

When model (1) was fitted to the test cricket first-innings scores, the Anderson–Darling test for normality indicated that the residuals were not normally distributed (p-value 0.000). However, by transforming the response variable using a logarithmic transformation, the normality assumption is not breached (p-value 0.165). Model (2) represents the subsequent multiplicative model:

$$\ln(s_{ijk}) = A + a_i - b_j + h_{ik} + \varepsilon_{ijk} \tag{2}$$

where $\Sigma_{i=1}^{9} a_i = 0$ and $\Sigma_{j=1}^{9} b_j = 0$. Applying the inverse transformation to the resultant parameters then gives estimates for the batting and bowling rating whose product is 1. These parameter estimates multiplied by 100 are provided in Table 6 and give the expected percentage of the average score $A = 276$ when that team is batting or bowling against an average opponent. For example, Australia's batting rating of 128 suggests that during the first innings it generally performed 28% better than average, whereas Zimbabwe's bowling rating of only 80 implies that its opponents scored $100/80 = 125\%$ of the average. Combining these results indicates that on a neutral ground against Zimbabwe we would expect Australia to score $128/80 = 160\%$ of the average. This figure would be adjusted if batting on home soil. Similarly, Zimbabwe against Australia would score $77/114 = 67\%$ of the average. Thus Australia would be expected to lead Zimbabwe on the first innings by a huge $160 - 67 = 97\%$ of the average first-innings score, or 268 runs.

Table 6. Team ratings for test cricket teams for the period 1997–2001

Team	*Batting rating*	*Bowling rating*	*Combined rating*
Australia	128	114	146
South Africa	118	120	142
New Zealand	104	100	104
Sri Lanka	104	99	102
Pakistan	105	94	99
West Indies	86	111	96
England	92	99	91
India	96	89	85
Zimbabwe	77	80	62

The common home advantage parameter was significant and estimated to be 112 (p-value 0.017), so the home team generally scored 12% more than expected on a neutral ground. This represents approximately 33 runs advantage. Again, there is no significant evidence to suggest that allowing for individual home advantages provides a significant improvement over the common home advantage model (p-value 0.967).

Although statistically not significant (p-value 0.167), when a parameter measuring the effect of batting first was included it was estimated at 0.93, which represents roughly 19 runs advantage to the team batting second.

The batting and bowling ratings can be combined multiplicatively to provide a measure of a team's overall strength, and they are also given in Table 6. Clearly, the results reinforce the outcomes that are summarized in Tables 1, 2 and 4, with Australia and South Africa rated substantially above average in both the batting and the bowling phases of test cricket. Notably, the West Indies has a significantly high bowling rating but is rated quite low with the bat. This is reflected in the fact that the West Indies won only 28% of its matches overall and could establish a first-innings lead only 33% of the time. This is an interesting result given that during the period of the study the West Indies had the game's premier batsman in Brian Lara playing for them. This reinforces the notion that the consistently successful teams tended to perform well in both batting and bowling. Since the chances of winning are increased after a first-innings lead has been established, solid first-innings performances in both batting and bowling tend to increase a team's chances of establishing a lead and subsequently increase its chances of winning. This is certainly the case with Australia and South Africa who won well over half their matches after establishing a first-innings lead. Comparisons with the one-day cricket ratings indicate that the rank order is preserved for the top two one-day teams. This suggests that the better-equipped teams tended to perform well in both forms of the game and underscores the recent dominance of Australia and South Africa in international cricket. Of the other teams, New Zealand appears to perform better at Test cricket whereas India performs poorer. As might be expected, the ratings for test cricket are more spread than those for one-day cricket, where restrictions that limit the balls bowled would even up results.

5. Modelling the factors affecting outcomes in test cricket

A multinomial logistic regression model is adopted to determine which factors associated with a team's first-innings performance most critically affect match outcomes. Using this approach we can examine, for example, whether the home or away team was more likely to win a test-match. A logistic regression model is employed because the response variable is categorical. The trichotomous response variable is the match outcome for the team batting first, which can be a win, draw or loss. It is also assumed that the response variable is ordinal because of the implicit order that is expressed in the match outcomes. The explanatory variables are the (signed) first-innings lead, a common home team advantage, the result of the toss and each team's batting superiority over the opposition's bowling (as measured by the team's batting rating divided by the opposition's bowling rating). If the cumulative probability of a win, draw and loss is denoted by γ_{ijw} for team i batting first in the second innings against team j, the outcome of a match is modelled as

$$\ln\left(\frac{\gamma_{ijw}}{1-\gamma_{ijw}}\right) = \beta_{0w} + \beta_1 h + \beta_2 x + \beta_3 \frac{a_i}{b_j} + \beta_4 \frac{a_j}{b_i} \tag{3}$$

where $w=0$ or $w=1$ for the respective cumulative probability of a win and draw, $h=1$ or $h=0$ if the home advantage rests with the team batting first or second and x is the (signed) first-innings lead of team i. Note that the (a_i/b_j)-term is the relative superiority of the batting team over the bowling team in the third innings and a_j/b_i is the relative superiority of the batting team over the bowling team in the final innings.

In originally fitting model (3) to the match outcomes we included a term for winning the toss but this proved not only insignificant but also in the wrong direction—i.e. winning the toss reduced a team's chance of winning the match. For this reason we have excluded the toss parameter from the model. This is an interesting outcome given that the established orthodoxy contends that winning the toss provides a team with a significant advantage, particularly if the wicket appears to favour its strengths. Captains tend to bat first when winning the toss and thus to avoid batting last on a possibly deteriorating wicket.

Assuming that relatively superior teams endeavour to capitalize on their strengths as well as to exploit the weaknesses of their opponents we examined the combined effect of winning the toss and team strength and included an interaction term to see whether relatively weaker teams were more prone to defeat when losing the toss. We found no evidence to support this hypothesis. The inability of captains to capitalize on winning the toss in one-day cricket was previously demonstrated by de Silva and Swartz (1997) and Clarke and Allsopp (2001).

The parameter estimates for model (3) are provided in Table 7. Application of the Pearson and deviance goodness-of-fit tests produces χ^2-values of 288.2 (p-value 0.617) and 245.1 (p-value 0.986) respectively, indicating that model (3) provides an adequate fit of the data. Clearly, the home team showed a strong tendency to win and the winning advantage that is gained by establishing a first-innings lead, though not surprising, is plainly evident. The latter point is highlighted in Table 2 where we find that most teams were inclined to win after

Table 7. Parameter estimates associated with the first innings of a test-match for the period 1997–2001

Parameter	*Term*	*Coefficient*	*p-value*
β_{00}	Intercept (win)	0.362	0.869
β_{01}	Intercept (win and draw)	1.986	0.367
β_1	Home	0.9257	0.011
β_2	Lead	0.006286	0.000
β_3	a_i/b_j	0.321	0.788
β_4	a_j/b_i	−2.485	0.053

establishing a first-innings lead. However, the model also produces a large advantage in batting second. A home team with bowling and batting ratings of 100, with no lead on the first innings over a similarly rated team, has about a 27% chance of winning if it bats first and 56% if it bats second. For the away team these reduce to 14% if it bats first and 35% if it bats second. This is clearly against current thinking but reinforces the trend that is shown in Table 3. There is also a clear indication that the probability of winning a test-match is highly influenced by the relative difference in strengths of the teams in the fourth innings but not the third innings. Although it was not surprising that the rating differential in the last innings is bordering on significant it was unexpected that the rating differential in the penultimate innings was insignificant. This suggests that teams are in the best position to force a win when they can exploit their superiority, whether that be batting or bowling, in the final innings. This may explain why the better teams are good at both bowling and batting—they are always in a position of strength in the final innings. It may also explain the low percentage of draws for the West Indies—its excellent bowling rating but weak batting rating means that either the West Indies or its opponents always have the upper hand in the final innings.

If we further characterize the overall strength of a team as the combined effect of its first-innings batting and bowling abilities we can restate model (3) as

$$\ln\left(\frac{\gamma_{ijw}}{1-\gamma_{ijw}}\right)=\beta_{0w}+\beta_1 h+\beta_2 x+\beta_3\,\frac{c_i}{c_j} \tag{4}$$

where $c_i=a_i b_i$ are the combined batting and bowling ratings from Table 5 and c_i/c_j represents each team's combined batting and bowling superiority over that of its opposition. Table 8 provides the parameter estimates for the new model. Application of the Pearson and deviance goodness-of-fit tests produces χ^2-values of 284.9 (p-value 0.683) and 246.3 (p-value 0.985) respectively, indicating that model (4) provides an adequate fit of the data. In general this model results in conclusions and observations that are similar to those of the previous model with all parameters now significant. It is not surprising to find that the relative difference in the competing team's overall first-innings batting and bowling strength has been a contributing factor in the shaping of a win.

To analyse the effect of establishing a first-innings lead, from model (4) we have

$$\Pr(w=0|h,x,c_i/c_j)=\frac{\exp(\beta_{00}+\beta_2 h+\beta_3 x+\beta_4 c_i/c_j)}{1+\exp(\beta_{00}+\beta_2 h+\beta_3 x+\beta_4 c_i/c_j)}. \tag{5}$$

Assuming that the team that bats first is the home team and competing teams are equally rated we have $h=1$ and $c_i/c_j=1$. If we set the probability of winning at 0.5 and solve equation (5) for the average lead that the home team needed to establish to have a better than even chance of winning we obtain 157 runs. Thus the home team needed a lead of 157 runs to have a 50%

Table 8. Parameter estimates associated with the first innings of a test-match for the period 1997–2001

Parameter	*Term*	*Coefficient*	*p-value*
β_{00}	Intercept (win)	−3.1616	0.000
β_{01}	Intercept (win and draw)	−1.5437	0.022
β_2	Home	0.8575	0.017
β_3	Lead	0.006363	0.000
β_4	Combined rating differential	1.3051	0.019

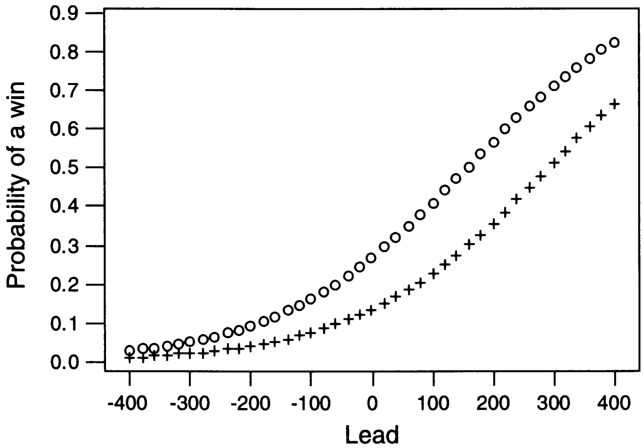

Fig. 1. Comparison of the estimated winning probabilities of the team batting first after establishing a first-innings lead for the home (○) and away (+) teams in test cricket for the period 1997–2001

chance of winning. Conversely, if the team batting first was the away team so that $h=0$ and $c_i/c_j=1$, then we find that the away team needs a lead of 292 runs to have a 50% chance of winning. This confirms the clear advantage that the home team had over the away team. Fig. 1 provides a comparison of the winning probabilities of the home and away teams for leads up to 400 runs. From Fig. 1, if two equally rated teams ended on the same score after the completion of the first innings, so that $x=0$, the point estimate probabilities of winning for the home and away teams were 0.27 and 0.14 respectively. However, in Section 4 we showed that the home side establishes an expected lead of approximately 33 runs on the first innings. Such a lead produces winning chances for the home and away sides of 0.31 and 0.11.

6. Conclusions

In analysing each innings of one-day cricket we are confronted with a resource inequity problem when the second-innings team wins. This arises because the first-innings team has exhausted available resources in the form of overs or wickets whereas when the second-innings team wins it has unused resources available to it. Consequently, to enable informed comparisons to be made between the first and second innings a projected 50-over score can be calculated that estimates the score that the second-innings team could make if it had access to unused resources. One-day scores can then be modelled to establish a team's relative batting and bowling abilities and subsequently to establish batting and bowling ratings. Similar modelling techniques can be applied to determine the batting and bowling ratings for teams in the first innings of a test-match. However, whereas an additive model adequately fits the one-day scores, a multiplicative model is a more appropriate fit for test scores. Clearly, Australia and South Africa have been dominant in both one-day and test cricket, which is reflected in their relatively high ratings in both forms of the game. A team's home advantage has also played a part, with the home team, on average, gaining a significant advantage (in runs) in both one-day cricket and the first innings of a test cricket match.

When modelling outcomes of matches in test cricket, not surprisingly, a team's first-innings lead is a very strong predictor of a winning result. The relatively stronger teams such as Australia and South Africa were more likely to establish a first-innings lead and thus showed a stronger inclination to win. There was a significant home advantage in test cricket. This came in two parts. The home team could generally establish a significant runs advantage over its opposition in the first innings, and it has been shown that the first-innings lead is a strong predictor of a win. This has provided the home team with a distinct winning edge. However, the home team also had a higher probability of turning any given first-innings advantage into a win.

Given that the first-innings ratings provide a measure of a team's relative strength over the duration of a test-match we can use the ratings to measure the relative difference in strength between teams during the third and final innings. The resulting model confirmed the descriptive statistics that it was an advantage to bat second. This challenges the established orthodoxy of generally electing to bat first when winning the toss and is probably the main reason why winning the toss is not helpful. It was also established that a team's superiority in the final innings was much more likely to produce a winning result than if its superiority over its opposition is exposed in the third innings.

Although the results generally confirm currently held beliefs on the relative strength of teams in one-day and test cricket, in some aspects the results are surprising. The model outcomes allow confirmation and quantification of well-known effects such as home advantage and first-innings lead but also provide opportunities to challenge aspects of cricket which have become folkloric, such as the effects that are attributed to winning the toss and the preferred batting order.

References

Allsopp, P. and Clarke, S. R. (2000) Methods for quantifying performances in one-day cricket. In *Proc. 5th Austr. Conf. Mathematics and Computers in Sport* (ed. G. Cohen and T. Langtry), pp. 28–36. Sydney: University of Technology Sydney.

Clarke, S. R. and Allsopp, P. (2001) Fair measures of performance: the World Cup of Cricket. *J. Oper. Res. Soc.*, **52**, 471–479.

Clarke, S. R. and Norman, J. M. (1995) Home ground advantage of individual clubs in English soccer. *Statistician*, **44**, 509–521.

Dixon, M. J. and Coles, S. G. (1997) Modelling association football scores and inefficiencies in the football betting market. *Appl. Statist.*, **46**, 265–280.

Duckworth, F. and Lewis, T. (1998) A fair method for resetting the target in interrupted one-day cricket matches. *J. Oper. Res. Soc.*, **49**, 220–227.

Harville, D. A. and Smith, M. H. (1994) The home-court advantage: how large is it, and does it vary from team to team? *Am. Statistn*, **48**, 22–28.

Lee, A. (1999) Modelling Rugby League data via bivariate negative binomial regression. *Austr. New Zeal. J. Statist.*, **41**, 141–152.

de Silva, B. M., Pond, G. R. and Swartz, T. B. (2001) Estimation of the magnitude of victory in one-day cricket. *Austr. New Zeal. J. Statist.*, **43**, 1369–1373.

de Silva, B. M. and Swartz, T. B. (1997) Winning the coin toss and the home team advantage in one-day international cricket matches. *New Zeal. Statist.*, **32**, 16–22.

Stefani, R. T. and Clarke, S. R. (1992) Predictions and home advantage for Australian rules football. *J. Appl. Statist.*, **19**, 251–261.