Forecasting test cricket match outcomes in play

Sohail Akhtar and Philip Scarf

Centre for Operations Management, Management Science and Statistics,
Salford Business School,
University of Salford,
Salford, Manchester M5 4WT, UK.
(email p.a.scarf@salford.ac.uk)

Salford Business School Working Paper Series

Paper no. 340/10

Forecasting test cricket match outcomes in play

Sohail Akhtar and Philip Scarf, University of Salford

Abstract

This paper forecasts match outcomes in test cricket in play, session by session. Match outcome probabilities at the start of each session are forecast using a sequence of multinomial logistic regression models. These probabilities can facilitate a team captain or management to consider an aggressive or defensive batting strategy for the coming session. We investigate how the outcome probabilities (of a win, draw, and a loss) vary session by session and how the covariate effects vary session by session. The covariates fall into two categories, pre-match effects (strengths of teams, a ground effect, home field advantage, outcome of the toss) and in-play effects (score or lead, overs-used, overs-remaining, run-rate, and wicket resources). Results indicate that lead has a small effect on the match outcome early on but it dominates later; pre-match team strengths, ground effect and home field advantage are important predictors of a win early on; wicket resources remaining are important throughout a match.

Keywords: Multinomial logistic regression; Strategy; Betting; Sport; Probability.

1. Introduction

Test cricket began in the late 1800s with matches that were not time-limited. In early matches, teams had an unlimited time to chase a target or to bowl out their opponents twice. Batting strategy was very simple in timeless matches; a team would seek to score as many runs as possible and to utilize all their batting resources. Only nine innings were ever declared in the ninety-nine timeless matches. The fixed-duration format was universally adopted in 1939, and in the present day, test matches are played over five consecutive days with three sessions in a day. To win a test match, broadly speaking, a team needs to bowl out their opponents twice within the time limit of five days. The five-day time limit necessitates strategic play particularly with respect to batting. A batting team needs to play each session according to a particular batting strategy in order to optimize the match outcome from their point of view. The optimal batting strategy in a session will depend upon the current match situation or match state. It is likewise for bowling. We use covariates to quantitatively summarise the match state at each stage, and we use the match state to forecast match outcome. We anticipate that these forecasts can then be used by team captains to guide strategic choice. They might also be used to rate player contributions.

Some work has been done on predicting match outcomes in test cricket: Brooks et al. (2002) use ordered probit with batting and bowling strengths, claiming to predict correctly 71% of outcomes. Scarf and Shi (2005) use logistic regression techniques to develop a model for match outcome probabilities given the end of third innings position. Scarf and Akhtar (2011) extend this work to

the end of first and second innings positions; their models are used to consider declaration strategy and the follow-on decision. Scarf et al. (2011) develop a model of match outcome given the match state at some point during the third innings, and use this model to consider batting strategy during the third innings. Our approach in this paper is new in that we model match outcome given the position at the start of each session. Match outcome probabilities are modelled using multinomial regression, with a win, draw, or loss response, and explanatory variables or covariates relating to match state, at the start of the each session. These covariates include the lead, used wicket resources of teams, run-rate, a home advantage factor, and surrogates for the state of the pitch (ground effect) and the pre-match strengths of teams. We also investigate how the covariate effects vary from session to session. We attempt to compare our results (graphically) with bookmakers' odds by means of examples. This is illustrated in the form of probability chart. Such charts may help management, players and fans to understand test cricket in the better way. Our study might be generalized to other sports. The fitting of a sequence of multinomial regression models also raises methodological questions about how best to fit such models that are at present beyond the scope of this paper.

There is other published work on test cricket that is related to ours here. Lenten (2008) analysed the decline of the frequency of draws in test cricket over the last fifteen years. Allsopp and Clarke (2004) have employed multinomial logistic regression to determine which the factors are associated with first innings batting performance. Crowe and Middeldrop (1996) studied the rates of leg before wicket dismissals between Australia and their visiting teams for test cricket series played in Australia between 1977 and 1994. Ringrose (2006) investigated leg before wicket dismissals in test cricket using generalized linear and mixed models. Clarke (1988) and Preston and Thomas (2000) have investigated optimal bating strategy in one-day matches. However, test matches are different because, in one-day matches, there is no notion of playing out the time remaining for a draw.

The remainder of the paper proceeds as follows. In the next section, the dataset on test match outcomes is briefly described, followed by a description of the modelling approach. The results of the model fitting are then presented, focussing on questions regarding which explanatory variables are important at each stage of the match, how strong is their influence, and forecasting ability, and how knowledge of the match outcome given the match position or state might be used to consider the batting and bowling strategy. We conclude with a discussion of the limitations of our work, and opportunities for further development.

2. Data Description

Data were obtained from the ESPNcricinfo website (ESPNcricinfo, 2010). Many forms of information are available on this website including, for example, ball-by-ball commentary, partnership scores, ground statistics and session by session match summaries. To obtain particular information, we have to use the ball-by-ball commentaries. An extract of the data used in our analysis is presented in table 1. The complete dataset (146 matches) relates to all the test matches in the period between November 2005 and March 2010, excluding those matches where the session by session data were not available and in which more than 90 overs were lost to poor weather. Session by session information is not generally available prior to November 2005.

The values of covariates of interest can be calculated from table 1, with the exception of team strengths and ground effect. To measure team strength, we use the International Cricket Council

(ICC, 2010) official ratings. These are published monthly. We use rating differences (RD), the difference in the ICC rating of the competing teams, as a covariate in the match outcome model. We also use win percentage difference (W%D), the difference in the win percentages of the competing teams, as a covariate and compare this to rating difference. We transform wickets into wicket resources. We define wicket resources for wicket i (i=1,…10). Partnership data for 197 matches collected by Scarf et al. (2011) was used to calculate the wicket resources (table 2). We use wicket recourses used (WR) as a covariate in our analysis. Lead is calculated session by session throughout the match with respect to the reference team, the team batting first in the match. Ground effect is the proportion of drawn matches on each ground over the period 1877 to 2010. This proportion for some pitches is shown in figure 1. The ground effect may vary over time, however, data were not available to us to assess this.

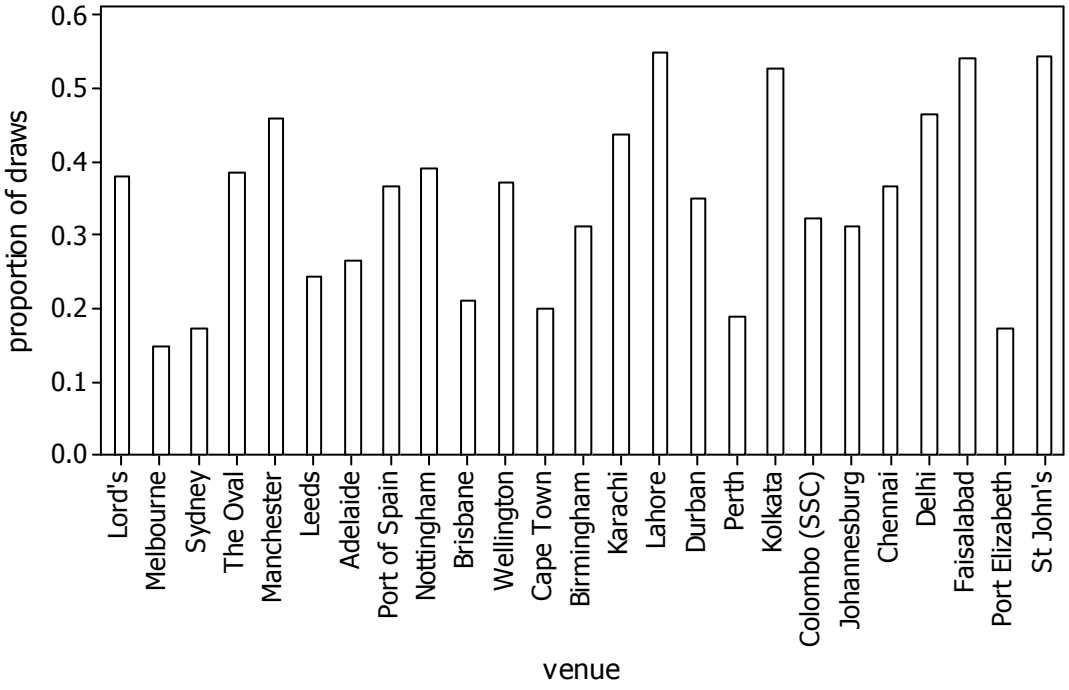


Figure 1.Proportion of drawn matches by venue for a selection of venues (1877 to 2010)

Table 1. Test match data (extract of 146 test matches, November 2005 to March 2010). Variables included: match date; teams, home team; runs scored in the each session; lead established at end of each session; total wicket resources used by reference team; total wicket resources used opponent team; win percentage difference; toss won by reference team; rating difference; team batting; estimated overs-remaining at the start of each session (calculated on basis of 90 overs per day (MCC, 2010)) and run-rate. Match result recorded as 1 (win), 0 (draw), -1 (loss) from point of view of reference team (batting first). Other variables not shown: wickets down at the end of each session; total wickets down for reference team and opponents; overs bowled in each session.

										At lunch-day 1							At tea-day 1						
Reference team	Opponents	Venue	Date	Percentage of drawn matches (G)	W%D	Home factor for ref. team(yes=1, no=0)	If ref. team win the toss then 1 else 0	Diff. b/w rating	Result	Reference team batting then 1 else 2	Runs in session	Total resources used (ref) WR_1	Total resources used (opp) WR_2	Run-rate	Overs-remaining in the match	Lead	Reference team batting then 1 else 2	Runs	Total resources used (ref) WR_1	Total resources used (opp) WR_2	Run-rate	Overs-remaining in the match	Lead
W	A	Hobart	17/11/2005	0.22	-50	0	1	-53	-1	1	49	24.0	0	1.88	424	49	1	124	62.6	0	2.25	395	124
W	A	Adelaide	25/11/2005	0.27	-50	0	1	-53	-1	1	71	37.5	0	2.63	423	71	1	194	51.1	0	3.46	394	194
I	SL	Delhi	10/12/2005	0.47	-2	1	1	13	1	1	68	24.0	0	3.78	432	68	1	164	37.5	0	3.15	398	164
A	SA	Perth	16/12/2005	0.19	25	1	1	28	0	1	96	11.1	0	3.69	424	96	1	175	37.5	0	3.30	397	175
I	SL	Ahmadabad	18/12/2005	0.50	-2	1	1	13	1	1	51	11.1	0	4.25	438	51	1	125	62.6	0	2.78	405	125
A	SA	Melbourne	26/12/2005	0.15	25	1	1	28	1	1	60	11.1	0	3.00	430	60	1	162	24	0	2.84	393	162
P	I	Faisalabad	21/01/2006	0.54	8	1	1	-12	0	1	137	24.0	0	3.81	414	137	1	248	51.1	0	3.88	399	248

Table 2. Wicket resources and cumulative wicket resources used as a function of wicket partnership number calculated from partnership scores in 197 test matches over the period of February 1998 to June 2004 (all innings); expressed as percentages.

Wicket	1	2	3	4	5	6	7	8	9	10
Resources	12.24	12.13	13.53	14.55	11.48	10.99	8.16	7.17	5.24	4.51
Cumulative wicket resources used	12.24	24.37	37.90	52.45	63.93	74.92	83.08	90.25	95.49	100.00

3. Modelling the match outcome

Scarf et al. (2005, 2011) and Scarf and Akhtar (2011) describe a multinomial regression model to explain match outcome probabilities given an end of the innings position. We use the same model to consider the multinomial response (win, draw, loss) as a function of the match position at the start of each session. With match outcome Y taking values (1, 0, -1) to denote a win, draw, and loss respectively, covariates denoted by X and taking a draw (0) as a reference category, this model assumes Y has a multinomial distribution, that is $Y \sim MN(p_1, p_0, p_{-1}; \sum p_i = 1)$, where p_1, p_0 and p_{-1} represent the probability of a win, a draw and a loss, with

$$\left.\begin{aligned} p_1 &= \exp(\alpha_1 + \beta_1^T X)/\{1 + \exp(\alpha_1 + \beta_1^T X) + \exp(\alpha_{-1} + \beta_{-1}^T X)\}, \\ p_0 &= 1/\{1 + \exp(\alpha_1 + \beta_1^T X) + \exp(\alpha_{-1} + \beta_{-1}^T X)\}, \\ p_{-1} &= \exp(\alpha_{-1} + \beta_{-1}^T X)/\{1 + \exp(\alpha_1 + \beta_1^T X) + \exp(\alpha_{-1} + \beta_{-1}^T X)\}. \end{aligned}\right\}$$

This model is the nominal multinomial logistic regression model.

The match outcome probabilities depend on the covariate vector X in different ways—through the coefficients β_1 and β_{-1} respectively. The model is equivalent in a sense to fitting two binary logistic regression models, the first for the win-draw probability comparison and the second for the lose-draw probability comparison. Test matches are played as a series of matches (except three triangular tournaments), so there is no points system in test cricket. The match outcome categories do not form a natural order. Therefore, it makes sense to consider match outcome as nominal. Other nominal multinomial regression models might be considered such as probit regression. We would however expect the results to be very similar. This is because the logistic and probit functions are very nearly linearly related over the interval $0.1 \leq p \leq 0.9$ (McCullagh and Nelder, 1989, p.109), and so it is difficult to distinguish the two functions based on goodness-of-fit. Ordinal (as opposed to nominal) probit regression has been used in football match prediction (e.g. Koning, 2000; Dobson and Goddard, 2003).

In all fifteen such models, one for each session, are fitted.

To assess model fit, we use the Akaike information criterion (AIC) (Sakamoto et al. 1986) and Nagelkerke R^2. This latter statistic is a modified form of Cox and Snell's pseudo R^2, and gives information about the explanatory power of the covariates in the model (Nagelkerke, 1991). Cox and Snell's Pseudo R^2 is given by

$$R^2_{CS} = 1 - \exp\{(-2(L_1 - L_0)/n\}$$

where L_1 represents the log-likelihood for the model with covariates, L_0 represents log-likelihood for the model with no covariates, and n represents number of observations. Nagelkerke's R^2 is given by

$$R^2 = \frac{1 - \exp\{(-2(L_1 - L_0)/n\}}{\{1 - \exp(2L_0)/n\}}$$

The maximum value of Cox and Snell's R^2 is $1 - \exp(2L_0)/n$ and therefore Nagelkerke's (modified) R^2 can vary from 0 to 1. Broadly speaking, Nagelkerke's R^2 for a generalized linear model is the percentage of variability in the outcome that is explained by the covariates in the model.

Rather than fit the sequence of regression models independently, one might consider sequential logistic regression in the manner of Elisheva et al. (2000). Here, covariates in model i in the sequence that applies at time t_i are the linear predictor or score from model i-1 that applies at time t_{i-1}- plus covariates that relate to new information that has arisen between t_{i-1} and t_i. This sequential approach is particularly useful when there are a large number of candidate covariates. Using the simpler approach, model selection can become a problem for models later in the sequence. Furthermore, using the score from a previous model as a covariate in a subsequent model implies that covariate effects in early models are still extant in later models but with a down-weighted effect. On the other hand, the direct explanatory power of particular covariates can be calculated over the sequence of model using the simpler approach. The fitting of sequential models is not straightforward because an errors-in-variables approach is required, since the linear predictor is a random variable. We therefore leave the question of the relative merits of the two approaches as an open question.

4. Model fitting results

4.1 Start of the match

Initially, we model match outcome at the start of a match. Outline model statistics for various fitted models are shown in table 3. We considered all possible factors affecting the outcome at the start of a match. Team strength (RD), ground effect (G) and home field advantage (H) were found to be important (based on AIC and Nagelkerke R^2). For team strength, we have used win percentage differences and the ICC rating differences and find that the rating differences have better explanatory power. This may be because the ICC rating takes account of result (win, draw, loss), along with the win margin, wickets and opponent rating. Winning the toss was also considered in the model fitting but was found to be unimportant. Ground effect (percentage drawn matches) can be thought of as a surrogate for the general quality of batting conditions. The playing conditions vary from ground to ground and country to country. For example, playing conditions in Lahore at the Gaddafi Stadium are quite different than in Leeds at Headingley. This is the reason that 55 percent of matches played at Gaddafi Stadium result in a draw while the figure for Headingley is 24 percent. We find that this factor plays a very important role in session by session analysis. We also note that Scarf and Akhtar (2011) and Scarf et al. (2011) failed to consider this as a covariate in their end of innings position models. Re-analysis of their declaration models indicates also that ground effect has significant implications for match outcome prediction given the end of first and

second innings positions. The prediction given the end of third innings position is not affected, because by the end of the third innings ground effect has broadly translated into target and overs-remaining.

Table 3. Start of first session position: results of fitting the multinomial logistic regression model to 146 test match outcomes for various sets of predictors: log-likelihood, AIC and Nagelkerke R^2. Covariates here are W%D, win percentage difference; RD, the ICC rating difference; H, home factor; G, ground effect, T; winning the toss; R-ratio, ratio of the ICC rating of teams.

Model	K	LL	AIC	Nag.R^2*(%)*
W%D	4	-135.688	279.376	27.09
RD	4	-135.063	278.126	27.83
R-ratio	4	-135.903	279.806	26.84
RD+H	6	-130.362	272.724	33.17
RD+H+G	**8**	**-126.766**	**269.532**	**37.03**
RD+H+**G**+T	10	-126.549	273.098	37.25

Using the highlighted model in table 3 (bold text), we can calculate the probability of win, draw and loss given the position at the start of a match. This will help team captains and management to consider their batting and bowling strategy for the first session. Estimates for the highlighted model are shown in table 4. Win and draw probabilities are presented in figure 2 for England when playing hypothetical matches against the rest of the current ICC member teams. It appears that team rating has a higher effect on the win probability than on the draw probability. The figure also shows that the England win probability at Lord's is considerably lower than at Headingley.

Table 4. Start of first session position: fitted parameter estimates for a start of the match minimum AIC logistic regression model (nominal) with covariates rating differences (RD), home factor (H), and ground effect (G), with standard errors and p-values, 146 test matches (November 2005 to March 2010).

		coefficient	s.e	p-value
win/draw (1/0)	Intercept	1.369	0.670	0.041
	rating difference (RD)	0.016	0.008	0.039
	home factor (H)	0.348	0.459	0.448
	ground effect (G)	-3.265	1.519	0.032
loss/draw (-1/0)	Intercept	1.967	0.680	0.004
	rating difference (RD)	-0.021	0.008	0.007
	home factor (H)	-1.044	0.511	0.041
	ground effect (G)	-3.761	1.603	0.019

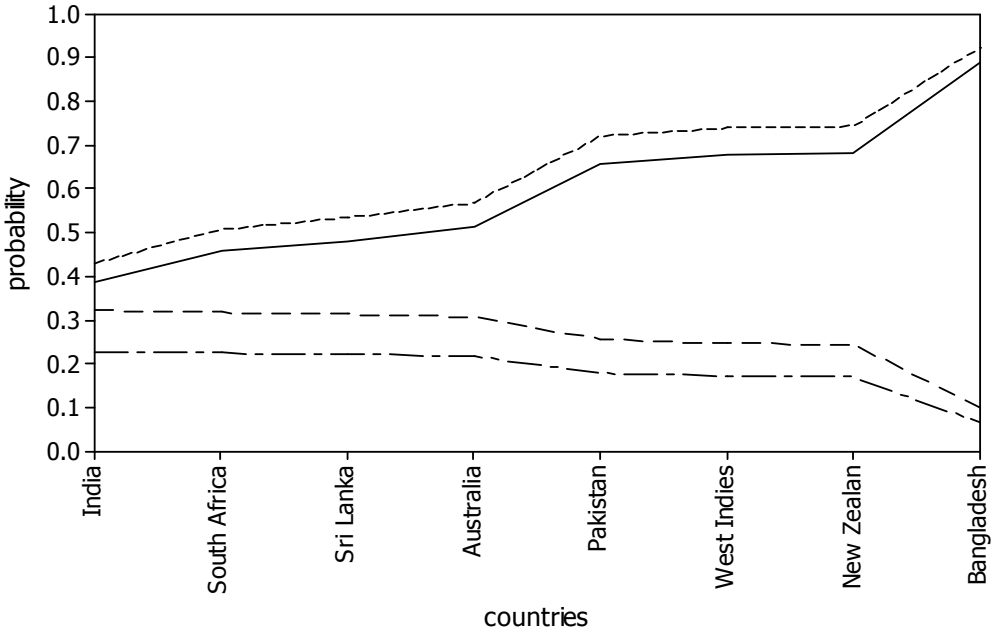


Figure 2. Win and draw probabilities for England cricket in hypothetical matches against various team at two grounds (Lord's and Headingley). ___ probability of win at Lords; ----- probability of win at Headingley; …… probability of draw at Lords; __.__. probability of draw at Headingley, based on October-2010 ICC ratings.

4.2 Models for subsequent sessions

Multinomial logistic regression models are considered at the start of each session. Only the best fitted models are given in tables 5. In the day-one models, lead, wicket resources used by the reference team, home factor, ground effect and team strength were found to be important (based on AIC and Nagalkerke R^2). In the day-two and day-three models, wicket resources used by the opponents was also found to be an important covariate. The explanatory power of pre-match strength, ground effect and home factor reduces with time. The home factor and ground effect becomes insignificant at tea on the fourth day, and the effect of pre-match strength is likewise by the start of the fifth day. The effects of these factors broadly translate into lead and used wicket resources. By the start of the fifth day, the effect of lead and wicket resources used by both the teams dominate as we would expect. The change over time of the explanatory power of team strength, ground effect and home factor is illustrated in figure 3. The explanatory power of an individual covariate is calculated as the difference between Nagelkerke's R^2 for the best fitting model, for the session, with and without the particular covariate.

In addition, we have used run-rate as a covariate to capture the present quality of the batting conditions but its effect was found to be insignificant. Note that we have ignored quadratic and higher order terms in the model fitting due to the limited number of observations in our data. Other factors that can influence the match result, such as the bowling and batting strengths, and the weather conditions are not quantified in our data. Captains would be expected to take into account these factors while considering batting and bowling strategy in each session.

Results in detail for a selection of the models are presented in the following sections.

Table 5. Results of fitting the multinomial logistic regression model to 146 test match outcomes for various sets of predictors: Nagelkerke R^2, % of correct match outcome predictions (within sample); explanatory power of ground effect, rating difference, and home factor. Covariates here are L, lead of reference team; RD, the ICC rating difference; H, home factor; G, ground effect; WR_1, used wicket recourses by reference team; WR_2, used wicket recourses by opponents.

Day	Session	Model	Nag. R^2(%)	Correct predictions %	Exp. power G (%)	Exp. power RD (%)	Exp. power H (%)	Number of matches
Day 1	Start of match	RD+H+G	37.0	59.6	3.9	24.3	5.2	146
	At lunch	RD+H+L+G+WR_1	43.8	63.7	3.9	14.1	5.4	146
	At tea	RD+H+L+G+WR_1	50.0	66.4	3.8	15.3	4.4	146
Day 2	Start of day	RD+H+L+G+WR_1	55.8	68.5	2.9	12.3	3.3	146
	At lunch	RD+H+L+G+WR_1+WR_2	57.4	71.2	3.3	10.6	2.5	146
	At tea	RD+H+L+G+WR_1+WR_2	62.6	75.3	3.0	7.9	2.2	146
Day 3	Start of day	RD+H+L+G+WR_1+WR_2	65.0	78.1	3.3	6.9	2.1	146
	At lunch	RD+H+L+G+WR_1+WR_2	67.8	75.3	3.6	4.6	2.3	146
	At tea	RD+H+L+G+WR_1+WR_2	73.2	80.1	2.9	4.4	1.8	146
Day 4	Start of day	RD+H+L+G+WR_1+WR_2	77.3	78.3	2.1	2.2	1.7	143
	At lunch	RD+H+L+G+WR_1+WR_2	74.9	80.2	1.6	1.7	1.2	131
	At tea	RD+L+WR_1+WR_2	76.8	79.5	0.9	1.7	0.9	122
Day 5	Start of day	L+WR_1+WR_2	79.9	81.1	0.8	0.5	0.6	111
	At lunch	L+WR_1+WR_2	83.7	86.8	0.2	0.6	0.2	91
	At tea	L+WR_1+WR_2	95.2	96.1	--	--	--	76

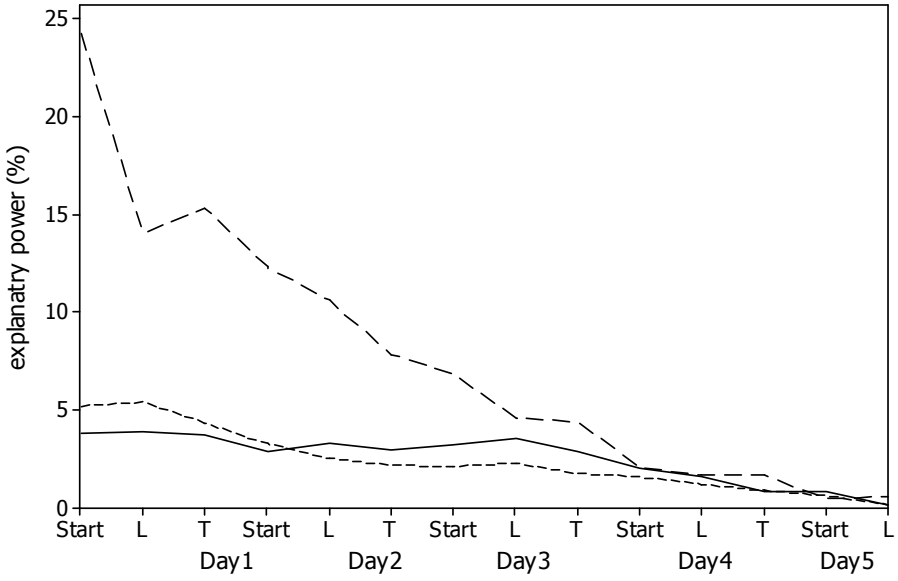


Figure 3. Session by session, the explanatory power of rating difference _ _ _ , home factor ------, and ground effect ___ .

4.3 Start of second day

Nominal multinomial logistic regression models considered at the start of day-two position are given in table 6. Estimates for the highlighted model is in table 8. At the start of the second day, lead, team strength, home factor, and wicket resources used by the reference team were found to be important (based on AIC and Nagelkerke R^2). The concordance table (table 7) shows that the model correctly predicts 68.5 % of match outcomes. This then shows that to an extent the model provides good prediction at this stage of the match. These prediction accuracy measures are based on in-sample forecasts so they should be treated with caution. We felt that there are insufficient matches in the data set to do out-of-sample forecasts. On the other hand, we might argue that we are not over-fitting models given that covariates sets for the best models are as we might have chosen *a* priori. Win, draw and loss probabilities for specified covariate values (for lead, team strength, home factor, and wicket resources used by the reference team) are shown in tables 18 and 19 for various values of lead and wicket resources.

Table 6. Start of day-two: results of model fitting the multinomial logistic regression model to 146 test match outcomes for various sets of predictors: log-likelihood, AIC and Nagelkerke R^2. Covariates here are L, lead; RR, average run-rate of first two sessions; WR_1, used wicket resources of reference team; H, home factor; G, ground effect.

Model	K	LL	AIC	Nag.R^2*(%)*
RD+H+L	8	-117.54	251.08	46.10
RD+H+G+L	10	-114.62	249.23	48.75
RD+H+G+L+WR_1	**12**	**-106.18**	**236.35**	**55.81**
RD+H+L+WR_1	10	-109.74	239.48	52.93
RD+H+G +L+WR_1+WR_2	14	-104.56	237.13	57.07
RD+H+G+L+WR_1+RR	14	-105.86	239.72	56.06

Table 7. Start of day-two: concordance table (cross-classification of observed and expected match outcomes) for validation data for nominal logistic regression the highlighted model in table 6.

	Predicted			
Observed	1	0	-1	Percent Correct
1	49	6	8	77.80%
0	16	12	5	36.40%
-1	9	2	39	78.00%
Overall Percentage	50.70%	13.70%	35.60%	68.50%

Table 8. Start of day-two: fitted parameter estimates for first innings minimum AIC logistic regression model (nominal) with covariates: lead (L), rating differences (RD), used wicket resources of ref. team (WR_1), and home factor (H), with standard errors and p-values, 146 test matches.

		coefficient	s.e	p-value
win/draw (1/0)	Intercept	2.071	1.319	0.116
	rating difference (RD)	0.018	0.008	0.020
	home factor (H)	0.437	0.477	0.360
	ground effect (G)	-3.300	1.530	0.031
	lead (L)	-0.006	0.004	0.100
	used wicket resources (WR_1)	0.015	0.009	0.122
loss/draw (-1/0)	Intercept	1.873	1.557	0.229
	rating difference (RD)	-0.018	0.009	0.042
	home factor (H)	-1.062	0.630	0.092
	ground effect (G)	-4.333	1.892	0.022
	lead (L)	-0.013	0.004	0.001
	used wicket resources (WR_1)	0.048	0.013	0.000

4.4 At lunch on third and fourth days

At lunch on the third and fourth days, the fitted models are given in tables 9 and 12 respectively. Maximum likelihood estimates for the highlighted (best) models are in tables 11 and 14. In both, position, lead, team strength, home factor, and used wicket resources of both team were found to be important (based on AIC and Nagelkerke R^2). The concordance table is also used to evaluate the predictive accuracy of the best fitted model (tables 10 and 13). Table 10 shows that the model correctly predicts 75.3 percent of the cases and table 13 shows 82.2 percent. Table 11 indicates that the win-draw probability ratio depends strongly on the opponents' used wicket resources, whereas the loss-draw probability ratio depends strongly on both current lead at the end of session and ground effect. On the other hand, table 14 shows that the loss-draw probability ratio depends strongly on both current lead and used wicket resources of the reference team. The win-draw probability ratio depends strongly on used wicket resources by opponents only.

Table 9. At lunch on day-three: results of the fitting multinomial logistic regression model to 146 test match outcomes for various sets of predictors: log-likelihood, AIC and Nagelkerke R^2. Covariates here are L, lead; RR, average run-rate of first day days; WR_1,used wicket resources of reference team; WR_2, used wicket resources of opponents; H, home factor and G, ground effect.

Model	K	LL	AIC	Nag.R^2*(%)*
RD +H+L+WR_1	10	-94.439	208.878	64.38
RD+L+WR_1+WR_2	10	-97.665	215.330	62.16
RD+H+L+WR_1+WR_2	12	-94.749	213.498	64.17
RD+G+L+WR_1+WR_2	12	-92.839	209.678	65.44
G+H+L+WR_1+WR_2	12	-96.250	216.500	63.14
RD+H+G+L+WR_1+WR_2	**14**	**-89.247**	**206.494**	**67.75**
RD+H+G+L+WR_1+WR_2+RR	16	-88.622	209.244	68.13

Table 10. At lunch on day-three: concordance table (cross-classification of observed and expected match outcomes) for validation data for nominal logistic regression the highlighted fitted model in table 8.

	Predicted			
Observed	1	0	-1	Percent Correct
1	50	6	7	79.40%
0	10	18	5	54.50%
-1	5	3	42	84.00%
Overall Percentage	44.50%	18.50%	37.00%	75.30%

Table 11. At lunch on day-three: fitted parameter estimates for first innings minimum AIC logistic regression model (nominal) with covariates: Lead (L), rating differences (RD), used wicket resources of ref. team (WR_1), used wicket resources of opponents (WR_2), home factor (H) and ground effect (G) with standard errors and p-values, 146 test matches (Nov. 2005 to March 2010).

		coefficient	s.e	p-value
win/draw (1/0)	Intercept	-2.560	1.941	0.187
	rating difference (RD)	0.009	0.009	0.274
	home factor (H)	0.667	0.517	0.197
	ground effect (G)	-4.279	1.674	0.011
	lead (L)	0.002	0.002	0.277
	used wicket resources (WR_1)	0.014	0.019	0.449
	used wicket resources (WR_2)	0.032	0.011	0.004
loss/draw (-1/0)	Intercept	-0.637	2.202	0.772
	rating difference (RD)	-0.018	0.010	0.067
	home factor (H)	-0.935	0.708	0.187
	ground effect (G)	-5.918	2.178	0.007
	lead (L)	-0.010	0.002	0.000
	used wicket resources (WR_1)	0.036	0.021	0.090
	used wicket resources (WR_2)	0.008	0.014	0.587

Table 12. At lunch on day-four: results of fitting the multinomial logistic regression model to 146 test match outcomes for various sets of predictors: log-likelihood, AIC and Nagelkerke R^2. Covariates here are L, lead; RR, average run-rate of first three days; WR_1,used wicket resources of reference team , WR_2, used wicket resources of opponents; H, home factor and G, ground effect

Model	K	LL	AIC	Nag.R^2*(%)*
RD+H+G+L+WR_1	12	-84.588	193.176	64.61
RD+H+G+L+WR_1+WR_2	**14**	**-69.035**	**166.070**	**74.92**
RD+H+G+L+WR_1+WR_2+RR	16	-68.247	168.494	75.38
RD+H+L+WR_1+WR_2	12	-71.756	167.512	73.29
H+G+L+WR1+WR_2	12	-71.854	167.708	73.23
RD+G+L+WR_1+WR_2	12	-71.084	166.168	73.70

Table 13. At lunch on day-four: concordance table (cross-classification of observed and expected match outcomes) for validation data for nominal logistic regression the highlighted fitted model in the table 12.

	Predicted			
Observed	1	0	-1	Percent Correct
1	53	3	3	89.80%
0	9	20	4	60.60%
-1	4	3	32	82.10%
Overall Percentage	50.40%	19.80%	29.80%	80.20%

Table 14. At lunch on day-four: fitted parameter estimates for first innings minimum AIC logistic regression model (nominal) with covariates lead (L), rating differences (RD), used wicket resources of reference team (WR_1), used wicket resources of opponents (WR_2), ground effect (G) and home factor (H) with standard errors and p-values, 146 test matches (Nov. 2005 to March 2010).

		Coefficient	s.e	p-value
win/draw (1/0)	Intercept	-9.293	2.551	0.000
	rating difference (RD)	0.005	0.009	0.563
	home factor (H)	0.413	0.609	0.497
	ground effect (G)	-2.951	1.727	0.088
	lead (L)	0.003	0.002	0.195
	used wicket resources (WR_1)	0.011	0.001	0.256
	used wicket resources (WR_2)	0.081	0.025	0.001
loss/draw (-1/0)	Intercept	-7.002	2.479	0.005
	rating difference (RD)	-0.011	0.011	0.277
	home factor (H)	-1.019	0.781	0.192
	ground effect (G)	-4.293	2.161	0.047
	lead (L)	-0.012	0.003	0.000
	used wicket resources (WR_1)	0.051	0.014	0.000
	used wicket resources (WR_2)	0.032	0.024	0.189

5. Forecasting match outcome

The objective of this paper is to provide a quantitative means for forecasting match outcomes in play. The quality of these forecasts is summarised in table 5. As the match progresses, the correct forecast ability of the models increases (from 59.6% at the start of the match to 96.1% at tea on the fifth day). It is also interesting to compare our model forecasts with forecasts based on bookmaker odds. Pre-match and in-play betting odds on test match outcomes are given by a number of bookmakers and betting sites. Such odds are available on Oddschecker (2010), but only in real-time. Betting odds histories were not available to us. We collected session by session odds offered on this web-site for six matches from 15 bookmakers. Three of these are presented as examples. The odds were converted into probabilities by accounting for the over-round, and then averaged over the bookmakers.

The session by session score cards of each example match are presented in tables 15 to 17. The session by session match outcome is forecast on the basis of the best fitting models in table 5. This

will provide a comparison of the work reported here with that of the bookmaker odds. This comparison allows us to make a number of points of interest in each example, below. In a final fourth example, we consider match outcome forecasts for the first test of the 2010-11 Ashes series which was in play at the time of writing.

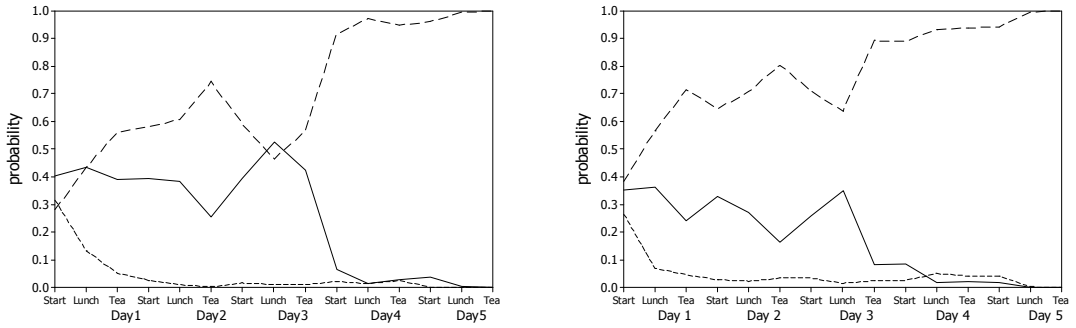
Example 1: The first example considers the match situation at lunch on day three of the second test match between Sri Lanka and India in 2010 at the Sinhalese Sports Club Ground, Colombo, Sri Lanka. According to our analysis, Sri Lanka's win and draw probabilities are 0.526 and 0.461 respectively (figure 4a). On other hand, bookmakers offered odds with implied win and draw probabilities 0.348 and 0.638 respectively (figure 4b). In our view, our result is more reasonable compared to bookmakers because at lunch on third day India had lost 3 wickets while still trailing by 469 runs in Sri Lanka.

Example 2: Next we consider the third test match is between Sri Lanka and India at the P.Saravanamuttu Stadium, Colombo Oval, Sri Lanka in 2010. Focussing on the start of the final day situation, according to our forecast, Sri Lanka win and draw probabilities are 0.64 and 0.10 (figure 5a). Bookmakers offered odds with implied win and draw probabilities 0.40 and 0.11 respectively (figure 5b). We believe that our forecasts are more reliable than those implied by the bookmaker odds because chasing a target of 203 runs in the 4th innings on the fifth day with 7 wickets in hand, away from home is quite difficult. Therefore we argue that Sri Lanka were in a better position to win than that implied by the bookmakers.

Example 3: Next we consider the third test match between England and Pakistan at the Oval, London in 2010 and the close of play on the third day. Our forecasts show England win and draw probabilities of 0.39 and 0.02 respectively (figure 6a). On the other hand, bookmaker odds imply England win and draw probabilities of 0.19 and 0.02 (figure 6b). Our forecast gives a much higher probability of a win than bookmakers because England were playing at home and at the time were a much higher rated team than Pakistan.

Table 15. Example 1: Scorecard for the 2nd test match between Sri Lanka (reference team) and India played at Sinhalese Sports Club Ground, Colombo, Sri Lanka. W_1 represents Sri Lankan total wickets down at end of session and W_2 represents Indian total wickets down at end of session.

Day	Day 1		Day 2			Day 3			Day 4			Day 5		
Session	At lunch	At tea	Start	At lunch	At tea	Start	At lunch	At tea	Start	At lunch	At tea	Start	At lunch	At tea
Lead	128	235	312	457	587	547	469	399	260	165	53	-27	-46	33
W_1	1	1	2	2	3	10	10	10	10	10	10	10	10	13
W_2	0	0	0	0	0	0	3	4	4	4	5	9	10	10

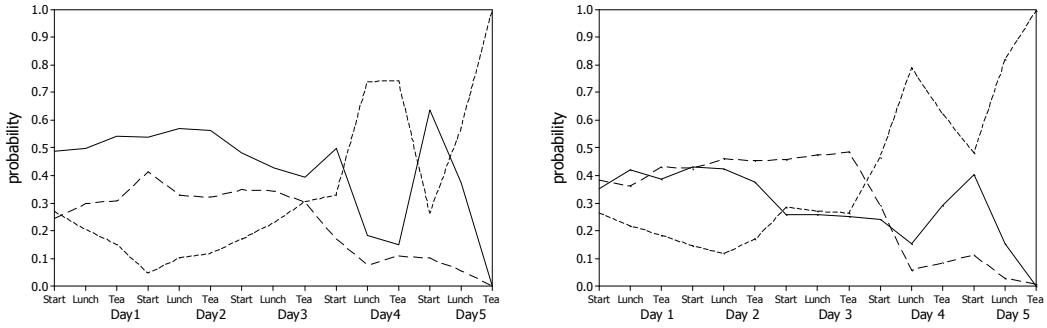


(a) (b)

Figure 4. a) Win probability forecast based on our models and (b) win probability based on bookmaker odds. ….. India win; ___Sri Lanka win; ------ draw.

Table 16. Example 2: Scorecard for the 3rd test match played at P Sara Oval, Colombo between Sri Lanka (ref. team) and India. W_1 represents Sri Lankan total wickets down at end of session and W_2 represents Indian total wickets down at end of session.

Day	Day 1		Day 2			Day 3			Day 4			Day 5		
Session	At lunch	At tea	Start	At lunch	At tea	Start	At lunch	At tea	Start	At lunch	At tea	Start	At lunch	At tea
Lead	102	194	293	369	397	245	143	47	34	132	214	203	111	-1
W_1	2	3	4	6	10	10	10	10	12	18	18	20	20	20
W_2	0	0	0	0	0	2	4	7	10	10	10	13	14	15

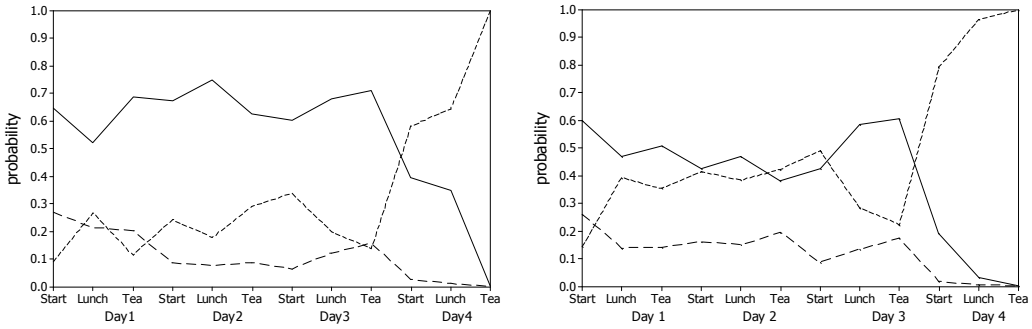


(a) (b)

Figure 5. a) Win probability forecast based on our models and (b) win probability based on bookmaker odds. ….. India win; ___Sri Lanka win; ------ draw.

Table 17. Example 1: Scored board of the 3rd match played at Kennington Oval, London between England (ref. team) and Pakistan. W_1 represents England total wickets down at end of session and W_2 represents Pakistan total wickets down at end of session.

Day	Day 1		Day 2			Day 3			Day 4		
Session	At lunch	At tea	Start	At lunch	At tea	Start	At lunch	At tea	Start	At lunch	At tea
Lead	70	175	185	122	18	-69	35	119	146	32	-1
W_1	5	5	10	10	10	11	12	13	19	20	20
W_2	0	0	1	4	5	10	10	10	10	13	16



(a) (b)

Figure 5. a) Win probability forecast based on our models and (b) win probability based on bookmaker odds. ….. Pakistan win; ___England win; ------ draw.

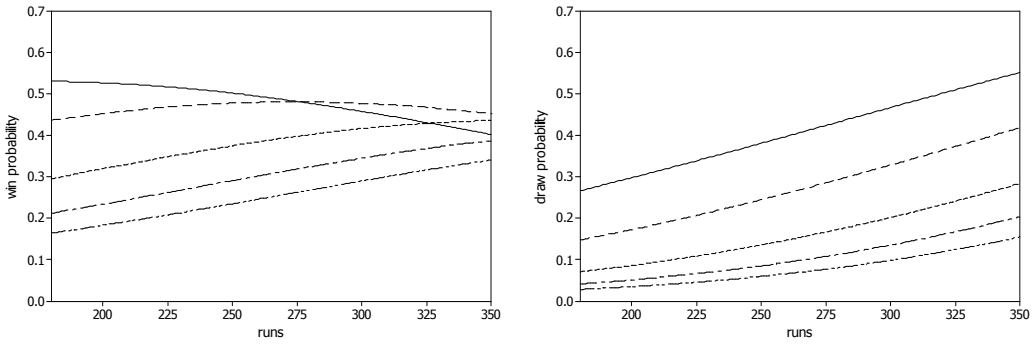
Example 4. The "Ashes" is a five-match series played between England and Australia, with a history of more than 100 years. At the time of writing, the 2010-11 Ashes series is in play in Australia. Australia have the home field advantage, but, England have a slightly higher ICC rating. This will be an interesting series for all stakeholders: fans, teams, and communication-media businesses. The first match of a series takes place in Brisbane and we forecast, using model estimates in table 8, match outcomes at start of second day position for various hypothetical positions (tables 18 and 19). From these we can deduce what are good and poor outcomes for the teams on the first day, for example. This kind of analysis might assist team captains and management to consider batting and bowling strategy. Broad effects are more evident in figure 7 (England batting first) and figure 8 (Australia batting first). The probability of a win increases with lead and with wickets remaining up to a point; if the lead is large and the wickets remaining is large then the win probability is lower as a draw is more likely. This is the case for both teams, although the win probability for Australia is higher overall given their home advantage.

Table 18. Win (W), draw (D) and loss (L) probabilities at start of second day position for Australia batting first as a function of lead established and used wicket resources used (WR_1); home factor, H=1; rating difference, RD= -2; ground effect (Brisbane), G=0.21 (proportion of draws).

Wicket down		200	230	260	290	320	350
2	W	0.677	0.650	0.619	0.584	0.545	0.504
	D	0.265	0.304	0.346	0.390	0.435	0.481
	L	0.058	0.045	0.035	0.027	0.020	0.015
3	W	0.686	0.669	0.646	0.617	0.584	0.547
	D	0.221	0.257	0.297	0.338	0.382	0.428
	L	0.093	0.073	0.057	0.044	0.034	0.026
4	W	0.677	0.672	0.659	0.640	0.614	0.583
	D	0.178	0.212	0.248	0.288	0.330	0.374
	L	0.145	0.116	0.092	0.072	0.056	0.043
5	W	0.650	0.657	0.656	0.646	0.629	0.605
	D	0.145	0.175	0.209	0.246	0.286	0.329
	L	0.205	0.168	0.135	0.108	0.085	0.066
6	W	0.600	0.621	0.633	0.636	0.631	0.618
	D	0.111	0.138	0.168	0.201	0.239	0.279
	L	0.289	0.242	0.199	0.162	0.130	0.103
7	W	0.554	0.584	0.605	0.618	0.622	0.617
	D	0.091	0.115	0.142	0.173	0.208	0.247
	L	0.355	0.302	0.253	0.209	0.170	0.136
8	W	0.515	0.549	0.577	0.597	0.608	0.610
	D	0.077	0.098	0.123	0.152	0.185	0.222
	L	0.408	0.352	0.299	0.250	0.206	0.167
9	W	0.474	0.513	0.546	0.573	0.590	0.599
	D	0.065	0.084	0.107	0.134	0.165	0.200
	L	0.460	0.403	0.347	0.294	0.245	0.201

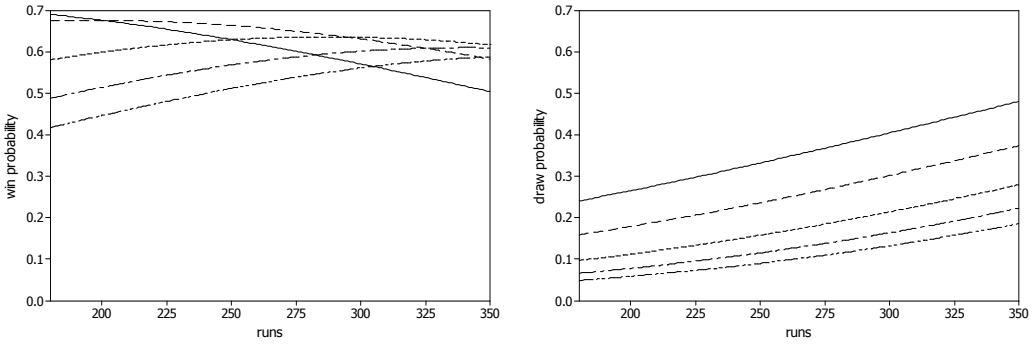
Table 19. Win (W), draw (D) and loss (L) probabilities at start of second day position for England batting first as a function of lead established and used wicket resources used (WR_1); home factor, H=0; rating difference, RD= 2; ground effect (Brisbane), G=0.21.

Wicket down		200	230	260	290	320	350
2	W	0.527	0.515	0.494	0.468	0.436	0.402
	D	0.297	0.347	0.398	0.450	0.501	0.552
	L	0.176	0.139	0.108	0.082	0.062	0.046
3	W	0.503	0.506	0.499	0.484	0.461	0.433
	D	0.233	0.280	0.330	0.382	0.435	0.488
	L	0.264	0.215	0.171	0.134	0.103	0.079
4	W	0.453	0.471	0.480	0.479	0.470	0.452
	D	0.172	0.214	0.260	0.311	0.364	0.418
	L	0.375	0.316	0.260	0.210	0.167	0.130
5	W	0.393	0.422	0.443	0.456	0.459	0.454
	D	0.126	0.162	0.203	0.250	0.301	0.355
	L	0.480	0.416	0.354	0.294	0.240	0.191
6	W	0.319	0.354	0.384	0.409	0.427	0.436
	D	0.085	0.113	0.147	0.187	0.233	0.284
	L	0.596	0.533	0.469	0.404	0.340	0.281
7	W	0.269	0.304	0.338	0.369	0.394	0.411
	D	0.064	0.086	0.114	0.149	0.190	0.237
	L	0.667	0.609	0.547	0.482	0.416	0.352
8	W	0.233	0.267	0.302	0.334	0.363	0.387
	D	0.050	0.069	0.093	0.123	0.160	0.203
	L	0.716	0.664	0.605	0.543	0.477	0.410
9	W	0.202	0.234	0.268	0.301	0.332	0.360
	D	0.040	0.055	0.075	0.101	0.134	0.173
	L	0.758	0.711	0.657	0.598	0.534	0.467



(a) (b)

Figure 7. Win (a) and draw (b) probabilities at start of second day position for England batting first as a function of lead established and used wicket resources used (WR_1): ___ 2 wickets down; ----- 4 wickets down …… 6 wickets down; __.__. 8 wickets down; _.._.. 10 wickets down; home factor, H=0; rating difference, RD= 2; ground effect (Brisbane), G=0.21.



(a) (b)

Figure 8. Win (a) and draw (b) probabilities at start of second day position for Australia batting first as a function of lead established and used wicket resources used (WR_1): ___ 2 wickets down; - ---- 4 wickets down; …… 6 wickets down; __.__. 8 wickets down; _.._.. 10 wickets down; home factor, H=1; rating difference, RD= -2; ground effect (Brisbane), G=0.21.

6. Discussion

The purpose of this paper is to model match outcome at the start of each session in test cricket. Looking at the start of session positions has two benefits. Firstly, some progress can be made with the quantitative analysis of the problem. Secondly, the models can guide the team captains and management with respect to batting and bowling strategy in each session. With $X(t)$ denoting the match state at time t, Y the match outcome, and S the strategy adopted in (t_0, t_1), the strategic decision problem can be stated as follows: first determine $\text{prob}(Y \mid X(t))$ (this is the model we develop in this paper); then determine $\text{prob}(X(t_1) \mid X(t_0), S)$ $(t_0<t_1)$; then compute $\text{prob}(Y \mid X(t_0), S) = \text{prob}(Y \mid X(t_1)) \times \text{prob}(X(t_1) \mid X(t_0), S)$ for various $S=s$ in (t_0, t_1); finally, choose S to "optimize" $\text{prob}(Y \mid X(t_0), S)$. This problem is straightforward to state. However, implementation is difficult because S is generally unobserved and so $\text{prob}(X(t_1) \mid X(t_0), S)$ is difficult to quantify. One solution is to explore different $X(t_1)$ scenarios (which are plausible given $X(t_0)$ and S) by considering $\text{prob}(Y \mid X(t_1))$ and subjective probability of the decision maker about the transition from $X(t_0)$ to $X(t_1)$ if he adopts strategy $S=s$ in the period (t_0, t_1). The model we develop in this paper facilitates this approach. From a less technical viewpoint, graphical presentation of probabilities $\text{prob}(Y \mid X(t))$ might also be of interest to the media and viewing public. Generalization of the methodology to other sports is possible in principal, although where a sport has few natural breaks (e.g. soccer, hockey, American football) the utility of the approach is questionable. An over-by-over analysis of the twenty-overs form of cricket might be amenable however.

Regarding particular characteristics of the model $\text{prob}(Y \mid X(t))$, our results show how the covariates that influence the match outcome vary from session to session. Early in the match, pre-match team strengths have a large effect. Home advantage is small while the ground effect (tendency for a draw at a particular ground) is large. These effects reduce as the match progresses. This is because these effects translate into lead and wicket resources as a match progress, so that later in a match, lead and wicket resources dominate. Of course other factors are influential, such as the weather condition, bowling and batting strength. One of the limitations of this work is that if

both the team lose wickets in the first day play then our models are not able to take account of wickets down for the team batting second. Data on more matches would be beneficial to overcome the problem.

As we would expect, the forecasting accuracy of the model improves as the match progresses. It would be interesting to carry out a study that compares the accuracy of the forecast from the model based on match state with those based on betting odds. We were not able to do this however as such match odds histories are not available to us. It would also be interesting to consider models fitted sequentially rather than independently, so that the linear predictor from the model at a particular stage is a covariate in the model at the following stage. This however would be another study. Finally, a potential application of the in-play forecasts we calculate would be the rating of batting, bowling and fielding contributions of players, based on the effect of player contributions on the predicted match outcome. This type of analysis of player ratings would be in the spirit of developed in soccer by Scarf and McHale (2005) and McHale et al. (2005).

References

Allsopp, P. E. & Clarke, S. R. (2004). Rating teams and analysing outcomes in One-day and Test cricket. *Journal of the Royal Statistical Society Series*, *A*, 167, 657-667

Brooks, R. D., Faff , R. W. & Sokulsky, D. (2002). An ordered response model of test cricket performance. *Applied Economics*, 34, 2353-2365.

Clarke, S. R. (1998). Test statistics. In: Bennett R (Eds). Statistics in Sport (1998). London: Arnold, pp 83–103.

Clarke, S. R. & Norman, J. M. (2003). Dynamic programming in cricket: choosing a night watchman, *Journal of the Operational Research Society*, 54, 838-845.

Crowe, S. M. & Middeldrop, 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.

Elisheva, S., Noya, G., Yana, Z. Dalit, B, & Benjamin, M. (2000). Sequential logistic models for 30 days mortality after CABG: Pre-operative, intra-operative and post-operative experience - The Israeli CABG study (ISCAB). *European Journal of Epidemiology*, 16, 543-555.

Dobson, S. & Goddard, J. (2003). Persistence in sequences of football match results: a Monte Carlo analysis. *European Journal of Operational Research*, 148, 247–256.

EPSNcricinfo (2010). Test match archives. http://www.stats.cricinfo.com/ci/content/records/307847.html/. Accessed on 10. October. 2010.

ICC (2010) Reliance Mobile Test Championship. http://icc-cricket.yahoo.net/match_zone/team_ranking.php, Accessed 28. October. 2010.

Koning, R. H. (2000). Balance in competition in Dutch soccer. *The Statistician*, 49, 419–431.

Lenten, J. A. (2008). Is the decline in the frequency of draws in test match cricket detrimental to the long form of the game? *Economic Papers: A journal of applied economics and policy*, 4, 364-380.

McCullagh, P. & Nelder, J. A. (1989). Generalized Linear Models. London: Chapman & Hall.

MCC. (2010). The laws of cricket. http://www.lords.org/laws-and-spirit/laws-of-cricket/laws/, accessed 01 August .2010

McHale I.G., Scarf P.A. & Folker D.E. (2010) On the development of a soccer player performance rating system for the English Premier League. *Submitted for publication.*

Nagelkerke N J D (1991). A note on a general definition of the coefficient of determination. *Biometrika*, 78, 691-692.

Oddschecker (2010).Oddschecker:http://www.oddschecker.com/cricket/ accessed July and August 2010

Preston, I. & Thomas, J. (2000). Batting strategy in limited overs cricket. *The Statistician*, 49, 95–106.

Ringrose, T. (2006). Neutral umpires and leg before wicket decisions in test cricket. *Journal of the Royal Statistical Society, series A*, 169, 903-911.

Sakamoto, Y., Ishiguro, M., & Kitigawa, G. (1986). Akaike Information Criterion Statistics. Tokyo: KTK Publishing House.

Scarf, P.A. & Akhtar, S. (2011). An analysis of strategy in the first three innings in test cricket: declaration and the follow-on. *Journal of Operational Research Society*, doi:10.1057/jors.2010.169.

Scarf, P.A. & McHale, I. (2005). Ranking football players. *Significance*, 2, 54-57.

Scarf, P. A. & Shi, X. (2005). Modelling match outcomes and decision support for setting a final innings target in test cricket: *IMA J. Management Mathematics*, 16, 161-178.

Scarf, P. A., Shi, X. & Akhtar, S. (2011). The distribution of runs scored and batting strategy in test cricket. *Journal of the Royal Statistical Society*, *series A: in press.*