DEVELOPING A TENNIS MODEL THAT REFLECTS OUTCOMES OF TENNIS MATCHES

Barnett, T., **Brown, A.** and Clarke, S.
Faculty of Life and Social Sciences, Swinburne University, Melbourne, VIC, Australia

ABSTRACT

Many tennis models that occur in the literature assume the probability of winning a point on service is constant. We show this assumption is invalid by forecasting outcomes of tennis matches played at the 2003 Australian Open. Revised models are formulated to better reflect the data. The revised models improve predictions overall, particularly for the length of matches and can be used for index betting. Suggestions on further improvements to the predictions are discussed.

KEY WORDS
tennis, sport, Markov chain, index betting

1. INTRODUCTION

Many tennis models that occur in the literature assume the probability of winning a point on service is constant (Schutz 1970, Carter and Crews 1974, Fischer 1980, Barnett and Clarke 2002). On the other hand, there are works in the literature to show that the assumption of players winning points on serve being *i.i.d.* does not hold. Jackson (1993), and Jackson and Mosurski (1997) show that psychological momentum does exist in tennis, and set up a "success-breeds-success" model for sets in a match, and find that this model provides a much better fit to the data, compared to an independence of sets model. Klaassen and Magnus (2001) test whether points in tennis are *i.i.d.* They show that winning the previous point has a positive effect on winning the current point, and at important points it is more difficult for the server to win the point than at less important points.

In this paper, an *i.i.d.* Markov Chain model is used to predict outcomes of tennis matches. The predictions indicate that the *i.i.d.* assumption may not hold since there are fewer games and sets actually played than predicted. A revised Markov chain model is then formulated for sets in a match that allows for players that are ahead on sets, to increase their probability of winning the set, compared to their probabilities of winning the first set. This is then followed by a revised model for games in a match that has an additive effect on the probability of the server winning a point. The revised models better reflect the data and the latter model is most useful for predicting lengths of matches, as demonstrated through index betting.

2. MARKOV CHAIN MODEL

2.1 MODELLING A GAME

A Markov chain model of a game for two players, A and B, is set up where the state of the game is the current point score (a, b), where both $a \geq 0$ and $b \geq 0$. With a constant probability p the state changes from (a, b) to $(a+1, b)$ and with probability $1-p$ it changes

from (a, b) to $(a, b+1)$. Therefore the probability $P(a, b)$ that player A wins the game when the point score is (a, b), is given by:

$P(a, b) = pP(a+1, b) + (1-p)P(a, b+1)$

where: p is the probability of player A winning a point.

The boundary values are $P(a, b) = 1$ if $a = 4$, $b \leq 2$, $P(a, b) = 0$ if $b = 4$, $a \leq 2$, $P(3, 3) = \frac{p^2}{p^2+(1-p)^2}$

Similarly, the mean number of points $M(a, b)$ remaining in the game at point score (a, b) is given by:

$M(a, b) = 1 + pM(a+1, b) + (1-p)M(a, b+1)$

The boundary values are $M(a, b) = 0$ if $b = 4$, $a \leq 2$ or $a = 4$, $b \leq 2$, $M(3, 3) = \frac{2}{p^2+(1-p)^2}$

Let $N(a, b|g, h)$ be the probability of reaching a point score (a, b) in a game from point score (g, h) for player A. The forward recurrence formulas are:

$N(a, b|g, h) = pN(a-1, b|g, h)$, for $a = 4$, $0 \leq b \leq 2$ or $b = 0$, $0 \leq a \leq 4$
$N(a, b|g, h) = (1-p)N(a, b-1|g, h)$, for $b = 4$, $0 \leq a \leq 2$ or $a = 0$, $0 \leq b \leq 4$
$N(a, b|g, h) = pN(a-1, b|g, h) + (1-p)N(a, b-1|g, h)$, for $1 \leq a \leq 3$, $1 \leq b \leq 3$

The boundary values are $N(a, b|g, h) = 1$ if $a = g$ and $b = h$.

2.2 MODELLING A SET

Let $P_A^{gs_T}(c, d)$ $\{P_A^{gs}(c, d)\}$ represent the conditional probabilities of player A winning a tiebreaker {advantage} set from game score (c, d) for player A serving. Let $P_B^{gs_T}(c, d)$ $\{P_B^{gs}(c, d)\}$ represent the conditional probabilities of player A winning a tiebreaker {advantage} set from game score (c, d) for player B serving.

The formulas below are for player A serving. Similar formulas apply for when player B is serving.

For a tiebreaker set:

$P_A^{gs_T}(c, d) = p_A^g P_B^{gs_T}(c+1, d) + (1-p_A^g)P_B^{gs_T}(c, d+1)$

The boundary values are $P_A^{gs_T}(c, d) = 1$ if $c = 6$, $0 \leq d \leq 4$ or $c = 7, d = 5$, $P_A^{gs_T}(c, d) = 0$ if $d = 6$, $0 \leq c \leq 4$ or $c = 5, d = 7$, $P_A^{gs_T}(6, 6) = p_A^{g_T}$.

where:
p_A^g and p_B^g represents the probability of player A and player B winning a game on serve respectively
$p_A^{g_T}$ represents the probability of player A winning a tiebreaker game

For an advantage set:

$P_A^{gs}(c, d) = p_A^g P_B^{gs}(c+1, d) + (1-p_A^g)P_B^{gs}(c, d+1)$

Boundary values: $P_A^{gs}(c,d) = 1$ if $c = 6$, $0 \leq d \leq 4$, $P_A^{gs}(c,d) = 0$ if $d = 6$, $0 \leq c \leq 4$, $P_A^{gs}(5,5) = \frac{p_A^g(1-p_B^g)}{p_A^g(1-p_B^g)+(1-p_A^g)p_B^g}$.

Recurrence formulas can be obtained for the mean number of games remaining in sets and the probability of reaching a game score in a set.

2.3 MODELLING A MATCH

Let $P^{sm}(e,f)$ represent the conditional probabilities of player A winning a best-of-5 set advantage match from set score (e,f).

The recurrence formula is represented by:

$$P^{sm}(e,f) = p^{s_T} P^{sm}(e+1,f) + (1-p^{s_T})P^{sm}(e,f+1)$$

Boundary values: $P^{sm}(e,f) = 1$ if $e = 3$, $f \leq 2$, $P^{sm}(e,f) = 0$ if $f = 3$, $e \leq 2$, $P^{sm}(2,2) = p^s$.

where:
p^{s_T} represents the probability of player A winning a tiebreaker set
p^s represents the probability of player A winning an advantage set

Recurrence formulas can be obtained for the mean number of sets remaining in a match and the probability of reaching a set score in a match.

3. MATCH PREDICTIONS

3.1 2003 AUSTRALIAN OPEN MEN'S PREDICTIONS

When two players, A and B, meet in a tournament, forecasting methods (Barnett and Clarke, 2005) are used to obtain estimates for the probability of each player winning a point on serve. These two parameters (each player winning a point on serve) are then used as input probabilities in the Markov Chain model to obtain match outcomes. We will compare the accuracy of the predictions to the actual outcomes for the 2003 men's Australian Open.

For a match between two players, the player who has greater than a 50% chance of winning was the predicted winner. Table 1 represents the percentage of matches correctly predicted for each round and shows that overall 72.4% of the matches were correctly predicted. Based on the ATP tour rankings only 68.0% were correctly predicted.

If p_i represents the probability for the predicted player for the i^{th} match, then the proportion of matches (P) expected to be predicted correctly and the variance (V) of the proportion can be calculated by:

$$P = \frac{\sum_i p_i}{n}$$

$$V = \frac{\sum_i p_i q_i}{n^2}$$

Table 1: Percentage of matches correctly predicted at the 2003 Australian Open

Round	Percentage correct(%)	No. of matches
1	78.1	64
2	62.5	32
3	68.8	16
4	75.0	8
5	75.0	4
6	50.0	2
7	100.0	1
Total	72.4	127

Table 2: Predicted and actual number of games and sets played at the 2003 Australian Open men's singles

	Games played	3 sets	4 sets	5 sets
Prediction	4737.7	41.1	42.7	34.2
Actual	4250.0	50.0	42.0	26.0

where:

$q_i = 1 - p_i$
n = total number of matches played in the tournament

Applying these equations gives values of $P = 0.753$ and $V = 0.0013$. The 95% confidence interval is represented by: $(0.753 - 1.96\sqrt{0.0013}, 0.753 + 1.96\sqrt{0.0013}) = (0.682, 0.824)$, which includes the value of 0.724.

Out of 127 scheduled matches for the 2003 Australian Open men's singles, only 118 were completed. For the other 9 matches, players had to withdraw prior to the match or retire injured during the match. Therefore, only the 118 completed matches were used for predicting the number of games and sets played. Table 2 gives the results. Overall, there were 487.7 fewer games played than predicted. This equates to $\frac{487.7}{118} = 4.13$ fewer games per match. Also, there were more 3 set matches played than predicted and fewer 5 set matches. This gives some indication that the *i.i.d.* model may need to be revised.

3.2 USING THE MODEL FOR GAMBLING

For head-to-head betting we will place a bet only when there is a positive overlay as represented by:

$$Overlay = [Our\ Probability \times Bookmakers\ Price] - 1$$

A method developed by Kelly, discussed in Haigh (1999), calculates the proportion of bankroll you should bet depending on your probability and the bookmaker's price and is represented below:

$$Proportion\ of\ bankroll\ to\ gamble = \frac{Overlay}{Bookmakers\ Price - 1}$$

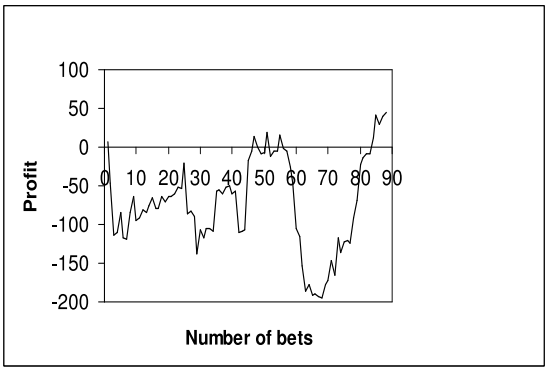
For example: Suppose player A was paying \$2.20 to win, and player B was paying \$1.65 to win. Suppose we predicted player B to win with probability 0.743.

In this situation we would bet on player B as given by a positive overlay:
$[0.743 \times 1.65] - 1 = 0.226$

$Proportion\ of\ bankroll\ to\ gamble = \frac{0.226}{1.65-1} = 0.348$

Figure 1 represents how we would have performed by adopting a constant Kelly system (fixed bankroll) of \$100 for the head-to-head matches played at the 2003 Australian Open. It can be observed that we would have suffered a \$195 loss by our 72nd bet but still ended up with a \$45 profit. This recovery came from round 3 (bet number 75) onwards, where at that point we were down \$147. By updating the parameters after each round by simple exponential smoothing (Bedford and Clarke, 2000) some important factors such as court surface, playing at a particular tournament, playing in a grand slam event and recent form would be included in the predictions.

Figure 1: Profit obtained from betting on head-to-head matches played at the 2003 Australian Open



Jackson (1994) outlines the operation of index betting with some examples in tennis through binomial-type models. The outcome of interest X is a random variable and for our situation is the number of games played in a tennis match. The betting firm offers an interval (a, b), known as the spread. The punter may choose to buy X at unit stake y, in which case receives $y(X - b)$ if $X > b$ or sell X at unit stake z, in which case receives $z(a - X)$ if $X < a$.

We will place a bet only when our predicted number of games is greater than b or less than a. For example if an index is (35,37), we would sell if our prediction is less than 35 games or buy if our prediction is greater than 37 games. We will use a very simple betting system, and that is to trade 10 units each time the outcome is favourable. Figure 2 represents how we would have gone by using our allocated betting strategy, for a profit of \$435. This was as high as \$480 but as low as -\$220. We also made \$420 from one match alone being the El Aynaoui versus Roddick match where a total of 83 games were played.

Without including this match we would have still made a profit of $60. Unlike head-to-head betting, there does not appear to be any advantage by betting from later rounds. We can generate a profit from the start of the tournament. Perhaps the bookmakers are not as proficient in estimating the number of games played in a match as they are with the probabilities of winning the match. The bookmakers are always trying to balance their books where possible so that they gain a proportion of the amount gambled each match regardless of the outcome. This implies that general public are unable to predict the number of games played in a match as well as probabilities of players winning. Figure 3 represents the results by subtracting an additional 4.13 games per match from our predictions. This gave a profit of $285, despite the fact that no money was bet on the El Aynaoui versus Roddick match, which made a $420 profit previously.

Figure 2: Profit obtained from index betting on matches played at the 2003 Australian Open

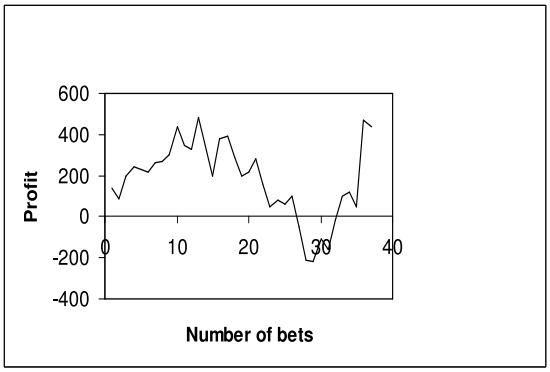
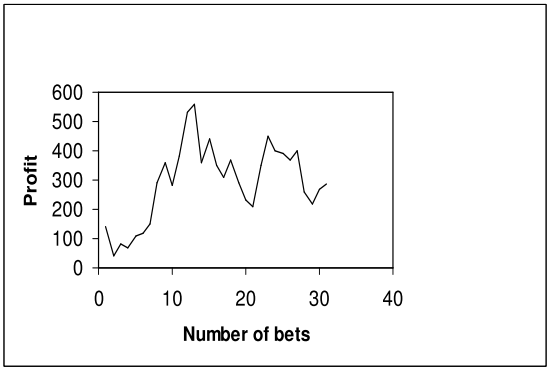


Figure 3: Profit obtained from index betting on matches played at the 2003 Australian Open by subtracting 4.13 games per match from our predictions



4. REVISED MARKOV CHAIN MODEL

4.1 PROBABILITIES OF REACHING SCORE LINES WITHIN AN ADVANTAGE MATCH

Suppose that if a player is ahead on sets, they can increase their probability of winning a set by α. Let $N^{sm}(e, f|k, l)$ be the probability of reaching a set score (e, f) in a match from set score (k, l). The forward recursion formulas become:

$N^{sm}(e,f|k,l) = p^{s_T} N^{sm}(e-1,f|k,l)$, for $(e,f) = (1,0)$
$N^{sm}(e,f|k,l) = (1-p^{s_T}) N^{sm}(e,f-1|k,l)$, for $(e,f) = (0,1)$
$N^{sm}(e,f|k,l) = p^{s} N^{sm}(e-1,f|k,l)$, for $(e,f) = (3,2)$
$N^{sm}(e,f|k,l) = (1-p^{s}) N^{sm}(e,f-1|k,l)$, for $(e,f) = (2,3)$
$N^{sm}(e,f|k,l) = (p^{s_T}+\alpha) N^{sm}(e-1,f|k,l)$, for $(e,f) = (3,0),(2,0)$ and $(3,1)$
$N^{sm}(e,f|k,l) = (1-p^{s_T}+\alpha) N^{sm}(e,f-1|k,l)$, for $(e,f) = (0,3),(0,2)$ and $(1,3)$
$N^{sm}(e,f|k,l) = (p^{s_T}-\alpha) N^{sm}(e-1,f|k,l) + (1-p^{s_T}) N^{sm}(e,f-1|k,l)$, for $(e,f) = (1,2)$
$N^{sm}(e,f|k,l) = p^{s_T} N^{sm}(e-1,f|k,l) + (1-p^{s_T}-\alpha) N^{sm}(e,f-1|k,l)$, for $(e,f) = (2,1)$
$N^{sm}(e,f|k,l) = (p^{s_T}-\alpha) N^{sm}(e-1,f|k,l) + (1-p^{s_T}-\alpha) N^{sm}(e,f-1|k,l)$, for $(e,f) = (1,1)$ and $(2,2)$

where: $0 \leq p^{s_T} + \alpha \leq 1$ and $0 \leq p^{s} + \alpha \leq 1$

The boundary values are $N^{sm}(e,f|k,l) = 1$ if $e = k$ and $f = l$.

Table 3 represents the probabilities of playing 3, 4 and 5 set matches when $\alpha = 0$ and 0.06, for different values of p_A and p_B. The probability of playing 3 sets is greater when $\alpha = 0.06$ compared to $\alpha = 0$, for all p_A and p_B. The probability of playing 5 sets is greater when $\alpha = 0$ compared to $\alpha = 0.06$, for all p_A and p_B.

Table 3: Distribution of the number of sets in an advantage match when $\alpha = 0$ and $\alpha = 0.06$

			$\alpha = 0$			$\alpha = 0.06$		
p_A	p_B	p^{s_T}	3 sets	4 sets	5 sets	3 sets	4 sets	5 sets
0.60	0.60	0.50	0.25	0.38	0.38	0.31	0.38	0.30
0.61	0.60	0.53	0.25	0.37	0.37	0.32	0.38	0.30
0.62	0.60	0.57	0.26	0.37	0.36	0.33	0.38	0.29
0.63	0.60	0.60	0.28	0.37	0.35	0.35	0.38	0.28
0.64	0.60	0.63	0.30	0.37	0.32	0.37	0.37	0.26
0.65	0.60	0.66	0.33	0.37	0.30	0.40	0.37	0.23
0.66	0.60	0.69	0.36	0.37	0.27	0.43	0.36	0.21
0.67	0.60	0.72	0.40	0.36	0.24	0.47	0.34	0.19
0.68	0.60	0.75	0.43	0.35	0.21	0.51	0.33	0.16
0.69	0.60	0.77	0.47	0.34	0.19	0.55	0.31	0.14
0.70	0.60	0.79	0.51	0.33	0.16	0.60	0.29	0.11

From our forecasting predictions in Subsection 3.1, it was noticed that on average the proportion of 3 set matches played are about 7% more than the model predicted and the proportion of 5 set matches are about 7% less than the model predicted, based on the assumption that the probability of players winning a point on serve are *i.i.d.* Notice from Table 3, the probability of playing 4 sets is about the same for both values of $\alpha = 0$ and 0.06, and the differences in probabilities for playing 3 sets is about 0.07 greater when $\alpha = 0.06$ compared to $\alpha = 0$, if $p^{s_T} \leq 0.75$. This was the reason $\alpha = 0.06$ has been chosen for the revised model.

4.2 CONDITIONAL PROBABILITIES OF WINNING A MATCH

The recurrence formulas are represented by:

$P^{sm}(e,f) = p^{s_T} P^{sm}(e+1,f) + (1-p^{s_T})P^{sm}(e,f+1)$, for $e = f$
$P^{sm}(e,f) = (p^{s_T}+\alpha)P^{sm}(e+1,f) + (1-p^{s_T}-\alpha)P^{sm}(e,f+1)$, for $e > f$
$P^{sm}(e,f) = (p^{s_T}-\alpha)P^{sm}(e+1,f) + (1-p^{s_T}+\alpha)P^{sm}(e,f+1)$, for $e < f$

Boundary values: $P^{sm}(e,f) = 1$ if $e = 3$, $f \leq 2$, $P^{sm}(e,f) = 0$ if $f = 3$, $e \leq 2$, $P^{sm}(2,2) = p^s$.

When $\alpha = 0$, the formulas reflect the Markov chain model presented in Subsection 2.3.

Table 4 represents the probabilities of player A winning an advantage match for $\alpha = 0$ and 0.06, for different values of p_A and p_B. Note once again that p^{s_T}, p^s and p^m represent the probabilities of player A winning a tiebreaker set, advantage set and an advantage match respectively. It can be observed that the probabilities remain essentially unaffected for all values of p_A and p_B by comparing the probabilities of winning the match when $\alpha = 0$ to $\alpha = 0.06$.

Table 4: Probabilities of player A winning an advantage match when $\alpha = 0$ and $\alpha = 0.06$

p_A	p_B	p^{s_T}	p^s	$p^m : \alpha = 0$	$p^m : \alpha = 0.06$
0.60	0.60	0.50	0.50	0.500	0.500
0.61	0.60	0.53	0.54	0.565	0.564
0.62	0.60	0.57	0.57	0.627	0.627
0.63	0.60	0.60	0.61	0.686	0.685
0.64	0.60	0.63	0.64	0.740	0.739
0.65	0.60	0.66	0.67	0.789	0.787
0.66	0.60	0.69	0.71	0.831	0.829
0.67	0.60	0.72	0.74	0.867	0.865
0.68	0.60	0.75	0.76	0.897	0.895
0.69	0.60	0.77	0.79	0.921	0.920
0.70	0.60	0.79	0.81	0.941	0.939

4.3 MEAN NUMBER OF SETS REMAINING IN A MATCH

The recurrence formulas are represented by:

$M^{sm}(e,f) = 1 + p^{s_T} M^{sm}(e+1,f) + (1-p^{s_T})M^{sm}(e,f+1)$, for $e = f$
$M^{sm}(e,f) = 1 + (p^{s_T}+\alpha)M^{sm}(e+1,f) + (1-p^{s_T}-\alpha)M^{sm}(e,f+1)$, for $e > f$
$M^{sm}(e,f) = 1 + (p^{s_T}-\alpha)M^{sm}(e+1,f) + (1-p^{s_T}+\alpha)M^{sm}(e,f+1)$, for $e < f$

Boundary values: $M^{sm}(e,f) = 0$ if $e = 3$, $f \leq 2$ or $f = 3$, $e \leq 2$, $M^{sm}(2,2) = 1$.

Table 5 represents the mean number of sets played in an advantage match M^{sm} for $\alpha = 0$ and 0.06, for different values of p_A and p_B. The mean number of sets played when $\alpha = 0.06$ is less than that when $\alpha = 0$ for all p_A and p_B.

Table 5: Mean number of sets played in an advantage match when $\alpha = 0$ and $\alpha = 0.06$

p_A	p_B	p^{s_T}	$M^{sm} : \alpha = 0$	$M^{sm} : \alpha = 0.06$
0.60	0.60	0.50	4.13	3.99
0.61	0.60	0.53	4.12	3.98
0.62	0.60	0.57	4.10	3.96
0.63	0.60	0.60	4.06	3.93
0.64	0.60	0.63	4.02	3.89
0.65	0.60	0.66	3.97	3.83
0.66	0.60	0.69	3.91	3.78
0.67	0.60	0.72	3.85	3.71
0.68	0.60	0.75	3.78	3.65
0.69	0.60	0.77	3.71	3.58
0.69	0.60	0.79	3.65	3.52

4.4 MEAN NUMBER OF GAMES IN A MATCH

We begin this subsection with the analysis of a tiebreaker match. The number of games in a tiebreaker match is finite, with the maximum of 65. However we are faced with the problem of relating the momentum for winning a set to the momentum for winning a game, or even winning a point.

A simple approach to dealing with this problem is to assume the momentum factor m has a linear form:

$$m = c_p k_p + c_g k_g + c_s k_s$$

where:
c_p, c_g, c_s are coefficients for terms of points, games and sets respectively.
k_p, k_g, k_s is the lead on the scoreboard in terms of points, games and sets respectively.

We can make further simplifying assumptions for the coefficients c_p, c_g, c_s, such that $c_s = 6c_g$ and $c_g = 4c_p$. The factors 6 and 4 come from the stopping rules for sets and games respectively. Putting $c_g = c > 0$ we then have:

$$m = c(\frac{1}{4}k_p + k_g + 6k_s)$$

where: $-4 < k_p < 4$ in a standard game, ($-7 < k_p < 7$ in a tiebreaker game, but this only occurs when the game scores are level at 6-all), $-6 < k_g < 6$ in a set, and $-3 < k_s < 3$ in a 5-set match. Combining all these inequalities we find that: $-25c < m < 25c$

Now if we model momentum as an additive effect on the probability of the server winning a point we require $p' = p + m$, where $0 < p + m < 1$. This puts theoretical limits on the values of c that can be assumed. It is possible to avoid such difficulties by following Jackson and Mosurski (1997) and express the momentum effect of leading on the scoreboard in terms of log odds:

$$\ln(\frac{p'}{q'}) = m \ln(\frac{p}{q})$$

This can be re-expressed as $p' = \frac{p}{p+qe^{-m}}$. It is easy to show that $p < p' < 1$ if the player is ahead, and $0 < p' < p$ if the player is behind. Whilst this theory appears nicer it is of no practical consequence, as the transformation is very close to linear in the main area of interest.

The linear model for the momentum factor enables us to calculate the probability of a player winning a point on serve in a tiebreaker game in a manner consistent with the probability of a point in a standard game. Furthermore wider consistency can be maintained with the probability of winning a game or a set. In particular the same linear model can be applied to an advantage set, even though the number of games is potentially infinite, because after 5-all has been reached the lead is at most one game until the end of the set. The probability of winning the advantage set will be consistent with the probability of winning the tiebreaker sets.

The model calculations have also been carried out and the results indicate, that under this momentum model, values of $0.0008 < c < 0.0032$ give a better fit to the data, when compared to an independence model ($c = 0$). *Nevertheless the linear model fails to fit the data on game difference and the pattern of sets simultaneously.* The model calculations have also been carried out with each player having his own base probability of winning a point on serve, as well as his own individual momentum factor to allow for temperament or other personality factors. However the model still fails to fit the data.

5. CONCLUSION

Using an Markov chain model to predict outcomes of tennis matches, where the probability of each player winning a point on serve is *i.i.d.*, overestimates the number of games and sets played in a match. In particular, from our forecasting predictions, it was noticed that on average the proportion of 3 set matches played are about 7% more than the model predicted and the proportion of 5 set matches are about 7% less than the model predicted. A revised Markov chain model is then formulated for sets in a match that allows for players that are ahead on sets, to increase their probability of winning the set, compared to their probabilities of winning the first set. This is then followed by a revised model for games in a match that has an additive effect on the probability of the server winning a point. The revised models are shown to better reflect the data, which could be useful for both punters and bookmakers, as demonstrated through index betting on the number of games played in a match.

Magnus and Klaassen (2001) have tested for independence of points from 4 years of Wimbledon point-by-point data. Further research for testing for independence, could involve analyzing Australian Open point-by-point data. This could involve finding a suitable momentum factor for individual players. In Subsection 4.4, we set up a revised model with the focus of estimating the number of games played in a match. This model could also be used for estimating the number of points played in a match, by choosing suitable values for the coefficients. As a further extension, this model can be modified to estimate the time duration of a match.

6. REFERENCES

Barnett, T. and Clarke, S.R. (2002) Using Microsoft Excel to model a tennis match, *In Proceedings of the 6M&CS*, G. Cohen and T. Langtry eds., 63-68.

Barnett, T. and Clarke, S.R. (2005) Combining player statistics to predict outcomes of tennis matches, *IMA Journal of Management Mathematics*, **16(2)**, 113-120.

Bedford, A.B. and Clarke, S.R. (2000) A comparison of the ATP rating with a smoothing method for match prediction, *In Proceedings of the 5M&CS*, G. Cohen and T. Langtry eds., 43-51.

Carter, W.H. and Crews, S.L. (1974) An Analysis of the Game of Tennis, *The American Statistician*, **28(4)**, 130-134.

Fischer, G. (1980) Exercise in probability and statistics, or the probability of winning at tennis, *Am. J. Phys.*, **48(1)**, 14-19.

Haigh, J. (1999) Taking Chances: winning with probability, *New York: Oxford University Press.*

Jackson, D.A. (1993) Independent trials are a model for disaster, *Appl. Statist.*, **42(1)**, 211-220.

Jackson, D.A. (1994) Index betting on sports, *The Statistician*, **43(2)**, 309-315.

Jackson, D.A. and Mosurski, K. (1997) Heavy Defeats in Tennis: Psychological Momentum or Random Effect?, *Chance*, **10(2)**, 27-33.

Klaassen, F.J.G.M. and Magnus, J.R. (2001) Are points in tennis independent and identically distributed? Evidence from a dynamic binary panel data model, *Journal of the American Statistical Association*, **96**, 500-509.

Schutz, R. (1970) A mathematical model for evaluating scoring systems with specific reference to tennis, *The Res. Quart.*, **41(4)**, 552-561.