Using official ratings to simulate major tennis tournaments. Stephen R. Clarke & D. Dyte.
International transactions in operational research 7(6) (2000): pp: 585-594(10).

Using Official Ratings to Simulate Major Tennis Tournaments

STEPHEN R. CLARKE and DAVID DYTE

Swinburne University of Technology, PO Box 218, Hawthorn, Vic. 3122,Australia

While the official Association of Tennis Professionals (ATP) computer tennis rankings are used to seed players in tournaments, they are not used to predict a player's chance of winning. However, since the rankings are derived from a points rating, an estimate of each player's chance in a head to head contest can be made from the difference in the players' rating points. Using a year's tournament results, a logistic regression model was fitted to the ATP ratings, to estimate the chance of winning as a function of the difference in rating points. Once the draw for a tournament is available, the resultant probabilities can be used in a simulation to estimate each player's chances of victory. The method was applied to the 1998 Men's Wimbledon, 1998 Men's US Open and the 1999 Men's Australian tennis championships.

Key words: sports, ranking, tennis, simulation, logistic regression.

INTRODUCTION

In many sports there is an officially accepted world ranking. Tennis has its Association of Tennis Professionals (ATP) World ranking, golf the Sony World ranking, and soccer its Fédération Internationale de Football Association (FIFA) ranking. While these create some measure of a player's or team's success and are sometimes used for seeding or prize money, they are not used in any predictive capacity. They are all based on some underlying points accumulation, and usually combine a mixture of qualitative and quantitative estimates.

The ATP tennis ranking is based on tournament and bonus points. Stefani (1997) gives details, but they alter each year. For 1998, a player gained tournament points depending on how far in a tournament he progressed, and the quality of a tournament. The maximum number of points available for any single tournament was 750 for winning a grand slam. Below that, tournaments were divided into Super 9, World Series, Challengers and Futures, with points allocated according to the prize money offered. At these last two levels, provision of hospitality for players also increased the number of points on offer. Bonus points were also awarded for defeating any player ranked 200 or higher, on a sliding scale. A maximum of 100 bonus points was available for defeating the number one ranked player in a major tournament. Points were summed from a maximum of 14 tournament results in the last 52 weeks. The points were updated at the end of each week, and published by the ATP on the WEB at www.atptour.com. In 1997 players' points ranged from 1 up to 5792 for Pete Sampras. This was made up of 4650 tournament points and 1142 bonus points. Of particular interest to players is the associated rankings, and a player's primary goal in tennis apart from winning a major tournament is to climb the rankings. These are used to seed players in the tournaments, to allow entry into some tournaments, and to allocate some prize money at the end of the year. However they are rarely used for prediction.

An exception to the ad hoc rating systems used in most sports is the Elo rating system used in chess (Elo, 1978). This purely quantitative rating system is based on exponential smoothing of a player's rating depending on the actual proportion of victory compared with that expected given the ratings of the opponents. The rating is not only used to rank current players, but to compare them with players from other eras. More importantly in the current context, there is a direct relationship in the difference in two players' ratings and their chance of victory - irrespective of the magnitude of the players' ratings. Thus a player who is ranked 100 points above an opponent will have a 64% percent chance of victory.

Various authors have suggested using exponential smoothing methods for rating in other sports, such as Strauss and Arnold (1987) for raquetball and Clarke (1994) for squash. One of the difficulties in ranking tennis players is the tournaments are played on different surfaces (grass, clay, synthetic etc), and may be indoors or outdoors. Most players have a favourite surface, and their performance level changes with different surfaces. For football, the fitting of linear models via least squares or other methods has provided a prediction method at least as good as the expert tipsters (Clarke, 1993; Harville, 1980; Stefani, 1980, 1987; Stefani & Clarke, 1992). These models also include a factor for home advantage, since it is known that teams perform better on their home grounds. Given unlimited resources, this method could be used to predict tennis results - fit a linear model which incorporates a player rating and a surface factor to a margin of victory, and use the resulting estimates for prediction. However the maintenance of such a system would require regular entry of tournament results. If predictions could be produced based on the ATP ratings, the problems of data collection would remain with the ATP. The aim of this study was to 'add value' to the official statistics, and produce a reasonable forecasting model. The strategy was to fit a model which gave the head to head chances of victory based on the players' ratings (or if possible, the difference in the ratings). The official ratings prior to a tournament were used to estimate each player's chance of victory over any other should they meet. A simple simulation based on the actual draw was then implemented to predict a player's chance of progression throughout the tournament or ultimate victory. This was updated as the tournament progressed.

DATA COLLECTION

To fit the model the results of all tournaments for a period, along with the official rating points at the start of each tournament were needed. While the official ATP site www.atptour.com updates the weekly rankings and ratings points each week (or after a fortnight, when a major tournament is being played), it does not archive the previous ratings. However another site (www.neiu.edu/~sgocha/tennis/tennis.htm) maintained by an interested tennis buff, had done this. A second site (gene.wins.uva.nl/~jellekok/tennis/) had the results of all tournaments. The data contained the usual problems of consistency: spelling of players' names was not consistent, and the format changed. The separate rankings data and the tournament data for 1997 were amalgamated into two text files, and a SAS program used to read the data, identify the inconsistencies, and finally merge the rankings and tournament information into a final single data set. Some detail was removed, such as the point score of tiebreakers, and any incomplete sets. For instance, in a match where a player retired during the second set trailing one set to love, only the first set was used in the modeling process. This resulted in a SAS data set containing four variables and 3003 observations (matches). For each match played in the year, this contained the rating points of the two players at the time of the match, and the result of the match in sets.

An initial check showed the ATP ranking proved moderately successful in correctly selecting the winner of individual matches for the four majors in 1997. In the Australian Open, the higher ranked player won 69.5% of matches. The corresponding figures for the French Open, US Open and Wimbledon were 60.9%, 62.5% and 64.1%. These figures were within the range reported by Stefani (1998) for predictions of other elite sports based on least squares. He quotes success rates of 63.4% for US pro football and 68% for Australian rules football. This gave some confidence that, at least in predicting the winner, the official rating system would not be too far behind a more complicated system we might have developed.

MODEL FITTING.

A logistic model was fitted to the data. If p is the probability of the higher rated player winning, then the logit of p or log of the odds ratio is

$$\ln\left(\frac{p}{1-p}\right) = a + bx$$

where x is the difference in ratings. For x = 0, p = 0.5, so a = 0. Thus

$$\ln\left(\frac{p}{1-p}\right) = bx$$

$$p = \frac{e^{bx}}{1+e^{bx}} = \frac{1}{1+e^{-bx}}$$

and $p \to 1$ as $x \to \infty$.

Because of the symmetry of the logistic curve with a=0, the order of participants in the data set does not affect the modeling results. That is, x = 200, with a set score of 2 to 1, has exactly the same effect in modeling as b = -200 and a set score of 1 to 2.

Initially the model was fitted using the probability of winning a match. While generally on the ATP tour matches are the best of three sets, grand slam tournaments are played as the best of five sets. Since the better player has a greater chance of winning a five set match, two models were needed, one for three sets and one for five set matches. However this meant the available data was severely reduced for five set matches, and it was this scenario we were particularly interested in. We decided to model the probability of winning a set. This had several advantages. It increased our data, removed problems of forfeited matches, and allowed the one model to be used for both three and five set match lengths. It also allowed our final simulation model to account for unfinished matches. Although the probability characteristics of tiebreaker and advantage sets differ, for the sake of simplicity both forms were treated equally in the modeling process. In our data, sets played under advantage rules, (i.e. the final set of five set matches in three of the majors), accounted for only 96 of 7566 sets, or 1.3%.

The advantages of modeling for sets rather than matches could equally apply to modeling for games rather than sets. However it is well documented that winning a game in men's tennis depends on whether a player is serving or receiving. Unfortunately the data set contained no information on which player served first in the match, or the number of service breaks each set. Both these would be necessary to reconstruct the number of service games each player won and lost.

The model was fitted with PROC GENMOD using SAS 6.12 to 7566 observations of the difference in the two players' rankings and a 0-1 variable indicating which player won the set. This resulted in a formula that would produce the probability of any ranked player winning a set against any other player. The value obtained for b implied that the number one player would win about 85% of sets against a newcomer with no points. This translates into a 94% chance of winning a three set match, and a 99% chance of winning a five set match.

SIMULATION OF WIMBLEDON.

The 1998 Wimbledon championship was the first major tournament to which the model was applied. Once the draw was made, the tournament was easily simulated. Given two players, the model gave the probability p of the higher ranked player winning a set. Assuming independence (see Pollard, 1983) the chance he will win a five set match in straight sets is p^3, in four sets $3p^3(1-p)$ and in five sets $6p^3(1-p)^2$. The chance he will lose is found by replacing p with $1-p$. While these formulas could be used to decide the winner in a head to head contest, it was decided to simulate at a set level. Thus a random number was generated to decide the winner of each set, and the match result tracked. This avoided the need for special coding to account for incomplete matches. The simulated winners were then advanced to the next round according to the actual draw, until a final winner was determined.

The program was written in qbasic, with results updated at the end of each day. 10 000 runs were used to generate estimates of the probabilities of each player winning the tournament or making the semi-final. Each day, the players with the highest estimated chances of winning were published on the Internet at www.swin.edu.au/sport/wim98/. Table 1 gives the pre-tournament output. In addition, specific matches of interest were chosen each day, and the probability of each of the six possible set scores, shown. An example, that of the final, is shown in Table 2.

The ultimate winner, Pete Sampras, was rated a 25% chance of winning prior to the tournament as shown in Table 1. His chance steadily increased up to 91% immediately prior to the final. On the other hand, the runner up Ivanisevic, did not appear in the daily tables of likely winners until after the fifth day, but this was still only part way through the second round. Table 3 gives the comparison of the model predictions and the observed winner and length of the 125 completed matches. Contrary to expectations, the higher ranked player won about the expected number of matches – 87 as against 82.7 predicted by the model. It was suggested prior to the simulation, that the unpredictability of grass and the substantial percentage of highly rated clay court players, would produce many upsets. However the actual number was within the statistical variation expected. However, in general matches were shorter than expected. Table 3 shows the number of 3, 4 and 5 set matches. For both the first round and the remainder of the tournament, the number of straight set wins was greater than expected. This remained true when subdivided into straight set wins to favourites and non-favourites.

Table 1
Pre-tournament estimated percentage chances of making the semi-finals and winning Wimbledon, 1998

Player	Semi-finals	Winner
Pete Sampras	53.9	24.6
Marcelo Rios	53.4	22.4
Petr Korda	49.3	14.4
Greg Rusedski	37.9	9.3
Carlos Moya	31.8	7.6
Pat Rafter	21.8	4.0
Yevgeny Kafelnikov	15.5	3.2
Alex Corretja	13.4	2.6
Jonas Bjorkman	17.1	2.5
Karol Kucera	11.6	1.8
Cedric Pioline	10.6	1.6
Richard Krajicek	9.6	1.4
Felix Mantilla	8.4	1.2

Table 2
Prediction of the 1998 Wimbledon final result

Winner	Score	Percentage chances
Sampras	3-0	43.3
Sampras	3-1	31.6
Sampras	3-2	15.4
Sampras	win	90.7
Ivanisevic	3-2	5.0
Ivanisevic	3-1	3.3
Ivanisevic	3-0	1.4
Ivanisevic	win	9.3

Table 3
Number of observed and expected results in completed matches for Wimbledon, 1998

Match Result	Whole Tournament Expected	Whole Tournament Actual	After Round One Expected	After Round One Actual
All matches				
3-0	37.4	55	19.5	26
3-1	46.2	47	23.2	29
3-2	41.4	23	20.3	8
Total	125	125	63	63
Favourites				
3-0	28.1	39	15.2	20
3-1	30.7	36	16.0	22
3-2	24.0	12	12.0	5
Total	82.7	87	43.3	47
Upsets				
3-0	9.4	16	4.3	6
3-1	15.6	11	7.2	7
3-2	17.4	11	8.3	3
Total	42.3	38	19.7	16

OTHER TOURNAMENTS

The method was also applied to the 1998 US Open and 1999 Australian Open and predictions published at www.swin.edu.au/sport/. Clearly the success of the model, as measured by the probability of success given in the early rounds to those players who ultimately progress through the tournament, depends on the degree to which results are in accordance with the seedings. Thus for the US Open, five of the quarter finalists and three of the semi-finalists were given in our pre-tournament list of 13 players with a better than 1% chance of winning. For the Australian Open, where the seeds tumbled rapidly, most quarter finalists did not make our list until late in the tournament. However the real advantage of the simulation is the interaction between ranking and draw difficulty. In Table 1 a player with a higher ranking generally has a higher chance of reaching the semi-finals and winning the tournament. This is not always so, particularly as the tournament progresses. Table 4 gives the estimated chance for the US Open of each quarter finalist making the semi-finals and winning the tournament. Carlos Moya, at that time ranked lower than Karol Kucera, is given three times the chance of winning the tournament. This is due to Kucera being in the more difficult half of the draw. He had to beat Pete Sampras (the 1997 runner up) and probably Pat Rafter (the 1997 winner) just to make the final. At the beginning of the tournament, when the spectre of meeting

Sampras and Rafter was a faint possibility, Kucera was given a better chance (3.1%) than Moya (2.0%) of winning. However as the tournament progressed, and Sampras and Rafter remained while rivals in Moya's half were eliminated, Moya's chance increased at a greater rate than Kucera's. We also see this effect in the relative chances of playing in the semi-finals and finals. Mark Phillippoussis was given a greater chance of making the semi-finals and final than both Kucera and Bjorkman, but less chance of winning the final. If either of the latter made the final then both Sampras and Rafter would have already been eliminated, and Kucera or Bjorkman would most likely play an easier opponent in the final than Phillippoussis would play. The total probability of all players from a particular part of the draw could be used as a measure of the draw difficulty. For example, Table 4 shows at the quarter final stage the top half (Sampras, Kucera, Rafter and Bjorkman) had a total probability of 74.5%, almost three times that of the bottom half. The simulation thus quantifies what the media often discusses.

Table 4
Day 9 estimated percentage chances of making the semi-finals and winning the US Open, 1998

Player	Semi-finals	Winner
Pete Sampras	73.0	36.9
Patrick Rafter	71.8	27.0
Carlos Moya	76.2	18.6
Karol Kucera	27.0	6.0
Jonas Bjorkman	28.3	4.6
Mark Philippoussis	55.4	3.4
Thomas Johansson	44.6	1.9
Magnus Larsson	23.8	1.6

DISCUSSION

Because of the timing, the actual ratings used prior to the tournament were a week out of date. However after the event, the simulation was rerun with the actual rating at the time the event began. This produced little change. One reason might be that many of the top players take a week off tournament play prior to slam events to practise.

While the ATP update the players' rating points after each tournament, it would be possible to update them during the tournament. Since opponents are always at the same stage of the tournament, tournament points earned would be the same, and hence would have no effect as the model works on the difference in rating. However the effect of bonus points can be considerable, particularly for players who defeat highly ranked players. Including these points as the tournament progressed might allow a lower ranked player to be given a rating more in keeping with current form. Another possibility would be to use an exponential smoothing type method during the actual tournament. This would have the effect of gradually replacing a rating based on 'reputation' with one based on current form as the tournament progressed. It would also allow the ease of a player's win to be taken into account, as such methods normally work on a margin of victory.

One problem that needs to be addressed is the under-prediction of the number of straight set matches. There may be several reasons for this. Although the ATP rating gives a measure of a player's average level throughout the year, on any particular day they may play significantly above or below this level. This introduces more variation than is present in our model and may produce more one sided matches than expected. A second possibility is the presence of a 'hot hand' effect. A player winning the first set gains confidence and thus has a higher probability of winning the next set. There are several possible methods for tackling this problem. One is to take a Bayesian approach, and alter the rating difference used for each set based on the actual set score in the match. This would require assumptions about the variation in tennis player's level of play from day to day. A second is to partition our original data into five subsets based on the set score (0-0, 0-1, 1-1, 1-2, or 2-2) and fit separate models to each data set. This would give the probability of a player winning the set conditional on the set score. One problem with this approach is the data available for model fitting are reduced, particularly for a 2-2 score line. Alternatively, the method used by Jackson (1993), where the odds on a player winning a set are increased by a constant factor for each set they are ahead, could be incorporated.

There is also the problem of alternative surfaces to be considered. It is well documented to successfully predict sporting results home advantage must be taken into account . While Holder and Nevill (1997) find little evidence of a home advantage in tennis, it is no accident that the semi-finals of the 1998 French open included three Spaniards and a Frenchman. The major reasons given for the existence of a home advantage are usually travel effects, crowd effects, and ground familiarity (Courneya & Carron, 1992). For tennis, there is probably little travel effect (even the home players have probably traveled from another international tournament), and some crowd effect. However the major effect is almost certainly ground familiarity, or court surface effects. It is well accepted in tennis that certain playing styles suit the different court surfaces, and some players are known as clay court or grass court specialists. Can this be taken into account?

The tournament organisers generally ignore this problem. The French Open is a good example. During his reign of several years as world number 1, Pete Sampras never won the French open, yet was consistently first seed. Wimbledon is the only tournament which departs from the ATP computer rankings in determining seedings. In this case, a simple method that can be applied to the top players is to reallocate points at the start of the tournament according to the seedings. So for example, a clay courter with 3456 points might be seeded tenth at Wimbledon by the seedings committee, between two other players with 2500 and 2600 points. For the purposes of the model, the clay courter could be allocated 2550 points. This has the disadvantage of introducing a subjective element, but would be easy to implement. It also cannot be used for unseeded players. A more complicated method would be to rate all tournaments according to surface, and calculate separate models for each.

One advantage of simulation is that the probability of virtually any event of interest can be estimated. While the main interest here was the probability of each player progressing to a certain stage, the chance of compound events is easily produced. The chance of two given players meeting, the chance the final will contain an Australian, the chance one player will progress further than another, can all be calculated. With the interest in the ATP rankings, the media often speculate on whether one player will pass another. In this case we are interested in the chance one player will gain at least a given number of (bonus and tournament) points more than another. This depends on whom they defeat as well as how far they progress, but is easily produced by the simulation. By interchanging players and re running the simulation at the start of the tournament, the advantages and disadvantages due to the draw could also be estimated.

Only men's tournaments have been simulated so far. The authors would like to redress this shortcoming, and produce some equality to our Web page by simulating the women's section of the major tournaments. Apart from equality issues, there has been speculation in the press that there are fewer upsets in the early rounds of women's tennis. This implies the pre tournament favourites produced by a model based on official ratings should be more successful. However we have not been able to find the necessary archived data to produce an analogous model for the Women's Tennis Association ratings.

CONCLUSION.

A simulation of Wimbledon based on official ATP rankings produced reasonable results. The simulation could be used to investigate difficulty of the draw, or assist in setting odds. For the two weeks of the tournament it provided an interesting discussion point, and was felt to be a worthwhile exercise. Some further work is needed to better allow for different surfaces, and better predict one-sided matches. The experiment was continued with the 1998 US Open and the 1999 Australian Open, and hopefully will continue for selected majors in the future.

While the current study used manual data entry and produced standard output, some work is currently being undertaken by colleagues to automate the process. An editor has been written in Java that allows simple updating of the current state of the tournament, including unfinished matches. It is hoped this will be further automated by logging the web sites set up by the tournament organizers. The simulator then performs a million simulations and loads the results into a database. A Java editor will then allow fans world wide to interrogate the database using a web browser. In this way sports followers can obtain up to date estimates of any aspect of any player’s chances in the tournament, rather than be restricted to a daily update of the statistics the authors think are of interest.

REFERENCES

Clarke, S.R., 1993. Computer forecasting of Australian Rules football for a daily newspaper. Journal of the Operational Research Society 44, 753-759.

Clarke, S.R., 1994. An adjustive rating system for tennis and squash players. In: de Mestre, N. (Ed.), Mathematics and Computers in Sport. Bond University, Gold Coast, Qld., pp. 43-50.

Courneya, K.S., Carron, A.V., 1992. The home advantage in sport competitions: A literature review. Journal of Sport & Exercise Psychology 14,13-27.

Elo, A.E., 1978. The rating of chess players, past and present. Batsford, London.

Harville, D.A., 1980. Predictions for National Football League games via linear-model methodology. Journal of the American Statistical Association 75, 516-524.

Holder, R.L., Nevill, A.M., 1997. Modelling performance at international tennis and golf tournaments: is there a home advantage? The Statistician, 551-559.

Jackson, D. A., 1993. Independent trials are a model for disaster. Applied Statistics 42, 211-220.

Pollard, G.H., 1983. An analysis of classical and tie-breaker tennis. Australian Journal of Statistics 25, 496-505.

Stefani, R.T., 1980. Improved least squares football, basketball and soccer predictions. IEEE Transactions on Systems, Man and Cybernetics 10, No. 2, 116-123.

Stefani, R.T., 1987. Applications of statistical methods to American football. Journal of Applied Statistics 14, 61-73.

Stefani, R.T., 1997. Survey of the major world sports rating systems. Journal of Applied Statistics 24, 635-646.

Stefani, R.T., Clarke S.R., 1992. Predictions and home advantage for Australian rules football. Journal of Applied Statistics 19, 251-261.

Stefani, R. T., 1998. Predicting Outcomes. In: Bennett, J.(Ed.)., Statistics in Sport. Arnold, London, pp. 249-273.

Strauss, D., Arnold B.C., 1987. The rating of players in racquetball tournaments. Journal of Applied Statistics 36, 163-173.

Clarke, S. R. and D. Dyte (2000). Using official ratings to simulate major tennis tournaments. *International Transactions in Operational Research* **7**: 585-594.