

Clarke, S. R. (2011). Rating non-elite tennis players using team doubles competition results.

This is the author's original draft of a work later published in *Journal of the Operational Research Society*, 62(7), 1385–1390.

Available from: <http://dx.doi.org/10.1057/jors.2010.75>

Copyright © 2011 Operational Research Society Ltd.

This is the author's version of the work, posted here with the permission of the publisher for your personal use. This version is the author's original draft and has not undergone peer review. It may vary substantially from the definitive version that appeared in the journal. You may also be able to access the published version from your library. The definitive version is available at <http://www.palgrave-journals.com/>.

Rating non-elite tennis players using team doubles competition results

Stephen R. Clarke

Faculty of Life and Social Sciences, Swinburne University of Technology, PO Box 218 Hawthorn, Victoria, AUSTRALIA: sclarke@swin.edu.au

Abstract. Statistical methods can be useful in rating non-elite tennis players. This paper shows how clubs can use simple optimization techniques to rate their club's players in doubles tennis competitions. Even though clubs lack all relevant information, the effects of home advantage, position played, partner and strength of team opposition can be taken into account and evaluated. The results from all competing clubs, available to the body organizing the competition, are used to rate all players in the competition, and validate the results that were obtained with limited information. We show a home ground advantage exists in non-elite doubles tennis. A simple exponential smoothing method of rating players is then tested, and shown to produce reasonable results.

Keywords: sports; tennis; regression; forecasting; exponential smoothing; home advantage

Introduction

Most statistical modeling in sport is applied to elite players and competitions. However the spirit of competition is often just as strong in lesser players, from juniors to veterans, from social to competition players. Rating players may be important to the individual wishing to measure improvement or achievement, selection committees wishing to choose between players for competition teams, and competition organizers wishing to group teams into sections of similar standard. This paper investigates the use of the results from a local veteran's doubles tennis competition to rate individual players.

Unlike golf, where each player can gain a handicap, there is no universal rating system for tennis. Golfers from different countries can compare their ability by quoting a single number, whereas even competition tennis players from the same town cannot compare standards. How does my playing number 3 in section 2 of Thursday afternoon over 50 veteran men's doubles competition compare with my son Rod playing number 1 in section 3 Wednesday night open competition?

(Stefani 1997; 1998) gives a summary of current methods of rating individuals and teams. He makes the distinction between accumulative and adjustive schemes. (Clarke, 1994) suggests an adjustive rating system for tennis which uses exponential smoothing similar to the (Elo, 1978) rating system used for chess. This adjusts up or down depending on whether players perform better or worse than suggested by their rating. Rating systems are often used for prediction, and clearly an adjustive scheme has an inherent prediction, as the actual performance is compared to that predicted by the relative ratings of the opponents. The relative prediction performance can be used to compare rating systems. (Bedford and Clarke, 2000) demonstrate that a simple adjustive system can be used to rate tennis professionals, and is marginally better than the ATP ratings in predicting likely winners. (Bedford, 2004) and (Ladds & Bedford, 2008) successfully apply a similar method to handball and badminton.

(Wright, 2009) suggests that more research should be directed at the non elite player. Tennis provides some unique opportunities for research, particularly at a non-elite level. Elite tennis is mainly an individual sport, with only a few team competitions (Davis Cup, Hopman Cup). But unlike most team sports, team competitions in tennis invariably consist of several individual matches, with the results accumulated in some way to provide a team result. Thus individual players can be successful while the team as a whole is not. Tennis does have 'doubles', a limited form of 'team' event where two players form a team, but at an elite level these teams would rarely change personnel, and would often be considered in ranking etc a single entity rather than two individuals.

The reverse is true of the vast majority of tennis played at a local level, where virtually all competitions are team events. A typical local competition might consist of a team of 4 or 6 players placed in order, with a match comprising various pairings playing the corresponding pair from the other team, with the team result determined by the team winning the most sets or games. Schedules are usually balanced with teams playing each opponent once on their home court and once away. Teams are usually grouped into sections containing

8 or 10 teams, with teams moving up or down a section the following season depending on performance or change of personnel.

The lack of a universal rating system for tennis players creates problems for club selection committees, which need to select players into teams. In the absence of any objective rating system, selections are based on personal preference or feeling, aided perhaps by some rudimentary collection of results. Similarly, competition organizers need to group teams of similar standard into sections. We ask the question “to what degree can standard quantitative methods assist in these processes of selection?”

Individual results

The data analysed in this study come from the Eastern Melbourne Mid-week Veterans Tennis Association. This doubles competition is for teams of 4 men aged over 50, and consists of about 5 sections of 8 teams. The team is entered on the card in some order 1, 2, 3 and 4 (usually, but not always, of decreasing ability) and a match consists of each of the 6 possible pairs (1&2, 3&4, 1&3, 2&4, 1&4, 2&3) playing the corresponding pair from the other team in a first to 8 (tiebreaker at 7 all) set. Thus each player plays 3 sets against his direct opponent, each of these including a different one of his team mates and his team mate’s direct opponent. With injuries and travel taking a toll, teams usually consist of 5 or 6 ‘regular’ players who rotate depending on availability, with emergencies called upon if necessary. The match winner is the team with the most sets (or in some seasons games), or games if sets are equal. The teams receive ladder points for winning the match, but also for each set won.

The results analysed here are for the author’s North Balwyn team in 2008 Section 2 Autumn season. The team consisted of 5 players EC, GM, SC, GB, and JZ. The team was very stable, no emergencies were used, and the playing order was generally in the order listed above. In common with many clubs, records are kept of individual player performances, and these may be used to assist selection. Individual totals such as those in Table 1 would usually be the most complicated statistics on a team season that most clubs would record.

Table 1 2008 Individual players’ season results

Player	Sets played	Sets won	Sets lost	Games up	Games up per set played
EC	33	21	12	36	1.1
GM	24	13	11	-7	-0.3
SC	27	18	9	44	1.6
GB	33	22	11	43	1.3
JZ	27	14	13	-6	-0.2

The most common statistic recorded appears to be games up or down over the season, as shown in the penultimate column of Table 1. While selection committees realize that other factors need to be taken into account, this is usually very haphazard. It is accepted that playing at 1 is more difficult than playing at 4, but the degree of difficulty is arguable. Other factors, such as the number of matches played, ability of partner, strength of opposition and home advantage are rarely allowed for. Here we at least normalize the figures to allow for the number of sets played to produce the final column. In this paper our usual dependent variable will be the number of games up per set played.

How much can be deduced from this table? While GM had the worst figures, playing at number 2 is clearly more difficult than at 3 or 4. Even JZ, with similar figures playing in the lowest position, could argue that he may have played with weaker players, or against stronger opposition. There are several other statistics available to selectors or organizers that could be used to shed light on performance – partner, position played, venue and strength of opponent. We look at a series of analyses that might be conducted if a club is interested in gaining more insights into player performance.

Club analysis of pairs results

The set results for each pairing over the season are shown in Table 2. Thus for example EC & GM won 8 games to 5 in round 1, but lost 3 games to 8 in round 8. Rounds 5 and 6 were a washout.

Clearly some pairs performed better than others and some simple analysis on the final column could be used to show this. For instance although EC always played in a higher position than GM, he performed at least 1.4 games per set better with each of his three partners. Using indicator variables to tag the individual players, we can now use PROC REG in SAS9.1 to fit games up as a linear function of the two players involved. On the 72 values in the body of the table, this gives a model significant at $p=0.07$ with an adjusted R square of 7.4%. But the parameters for the players in decreasing order are $SC=1.4$, $GB=1.2$, $EC = .8$, $JZ=-0.9$, $GM =-0.9$. An analysis using just the 10 averages in the final column of Table2 gives virtually the same figures. Note this is the same order that would be given by the final column of Table 1, so not much has been achieved by this more detailed analysis. However the results are at odds with the general acceptance in the team that EC is clearly the best player. It might still be argued that one pair or player played more games at home, or against stronger opponents. A more detailed analysis can quantify other effects.

Table 2 Games score in each set for each pair in each round

Players		Round											Games up	Games up per set	
		1	2	3	4	7	8	9	10	11	12	13			14
EC	GM	8-5	8-4			8-6	3-8		8-4	4-8			2-8	-2	-0.3
EC	SC	8-2			5-8	8-6		8-3		8-6	6-8	8-1	8-5	20	2.5
EC	GB	8-3	8-5		8-3	8-4	4-8	7-8	8-5		8-6	8-3	6-8	20	2.0
EC	JZ		8-7		8-3		2-8	8-7	5-8	8-4	5-8	7-8		-2	-0.3
GM	SC	8-7		4-8		4-8				8-2			8-7	0	0.0
GM	GB	7-8	8-6	8-6		8-1	4-8		8-6				4-8	4	0.6
GM	JZ		8-5	2-8			2-8		6-8	8-6				-9	-1.8
SC	GB	8-2		8-2	8-3	8-7		7-8			8-3	8-7	4-8	19	2.4
SC	JZ			3-8	8-2			4-8		8-2	6-8	8-4		5	0.8
GB	JZ		8-6	5-8	8-5		3-8	8-4	8-6		8-6	3-8		0	0.0

However the result sheet also records the venue (home or away), the opposing club, and the position the individuals played. Since each set contains two players, we define a new variable ‘set weakness’ as the sum of the playing order of the two players. Thus the set containing 1& 2 has a set weakness of 3, 1&3 a set weakness of 4, 1&4 and 2&3 a set weakness of 5, 2&4 6 and 3&4 7. These are clearly in the correct decreasing order of difficulty, with the only arguable point the equality of 1&4 and 2&3.

Since it is difficult for an individual club to get a measure of the ability of a person’s direct opponent on the day (this might be possible for an association, which has the records of all the clubs), we need a surrogate. At the end of the season final ladder details give a measure of the strength of the opposing club throughout the season. Since final ladders are published they are available to all clubs. A range of measures could be used – ladder points, matches won and lost, sets won and lost, games won and lost. A set can be won in a tiebreaker, so becomes more variable and less descriptive of the relative strengths than the actual game score. A score that is as close as possible of 8 games to7 becomes 1 set to love. This is repeated for a match as compared to sets. Hence it can be argued that the ladder measure that contains the most information is game percentage – $100 * \text{games won} / \text{games lost}$. Adding the final percentage of the opposing club (obtainable from the final published ladder) to the above data set allows a measure of the strength of the opposing club. Thus to the data of Table 2, we can append the variables home (1 or 0), weakness of set (1 to 7) and final ladder percentage of opposing team

This gives a highly significant model ($p=0.0004$) with an adjusted R square of 26%. Clearly there is still much unexplained variation, presumably variation in individual player form and the individual opponent strength, and of course random variation (the net cord at set point). The parameter estimates given in Table 3 are revealing. With the exception of GM they are in team playing order. Only EC showed a parameter estimate significantly different from zero. Since on average players in the section would be zero games up (one player’s win is another’s loss) this suggests EC should be in a higher section, with no significant

evidence the other players are misplaced. The set weakness is almost significant at the 10% level, with each position being 1.2 games per set more difficult. A little care is needed in interpreting this, as each increment in position only means an increase of 2 in total over 3 sets. Thus playing at number 1 is on average about 2.4 games per day, or 0.8 games per set more difficult than at number 2, and 2.4 games per set more difficult than at number 4. The home advantage is a surprisingly high value of 1.7 games per set. There is usually only local travel in this competition, little crowd support, and minor variation in court surrounds. In fact the home team has some duties (court maintenance and supply of afternoon tea) that might be expected to be a disadvantage to good play. The most surprising is strength of opposition. This is highly significant, and at -0.064 may seem small. However final ladder percentage varied from a low of 66 to 139. Thus the expected difference in playing against the top and bottom team is $(139-66) \cdot 0.064 = 4.7$ games a set, or about 14 games over the day. A swap between the two players rostered out against the top and bottom teams respectively could make a difference of 28 games, enough to wipe out most of the player differences in season totals shown in table 1. Yet the strength of the opposing team is virtually never discussed as a qualifying factor in players' statistics.

Individual comparisons show that EC is better than GM and JZ; GM, SC and GB are better than JZ; and all other comparisons are insignificant. This is generally consistent with the team's personal feelings.

Although we have used a statistical package to produce the estimates and conduct the statistical tests, the ratings themselves (ie the parameters) are easily produced on Microsoft Excel by using Solver to minimize the sums of squares of the errors. Such a method is within the capabilities of most clubs.

Note that while this method is useful for comparing players within a team, unless some players have played in both teams, it does not help to compare players between teams. If one player is shown to be significantly worse than the average for that section, while a player in a lower team is significantly better than the average for that section, that could be used as an argument for team changes.

Table 3 Parameter estimates for effects on games up per set

Variable	Parameter Estimate	Standard Error	t Value	Pr> t
EC	2.5	1.2	2.0	0.046
GM	-0.1	1.8	-0.1	0.956
SC	1.0	1.9	0.6	0.585
GB	0.1	2.5	0.1	0.959
JZ	-3.1	2.5	-1.2	0.221
Set weakness	1.2	0.7	1.7	0.102
oppct	-0.1	0.0	-3.5	0.001
home	1.7	0.8	2.0	0.050

Association Analysis

The largest source of unexplained variation that remains in the above analysis is the ability of the individual opponent. In our team, clearly EC is the best player, and a case can be made for him playing in a higher section. When he is rostered out of the team, the opposing number 1 has an easier time than when he is in. Our players' results are similarly affected by the availability of opposing players. However the association has the records of all matches, and a larger regression with all players included as effects is feasible. The association running the competition has available the results for all clubs, and they have provided the records of all matches. The Section 2 data comprised 52 players and 276 sets. This shows on average clubs use nearly 7 players over the season, almost double the minimum number of 4 required. These data were used in a larger regression involving a constant home advantage and separate ratings for each individual player.

$$\text{Set margin} = \text{home advantage} + \text{homeplayerA} + \text{homeplayerB} - \text{awayplayerA} - \text{awayplayerB} + \text{error}$$

Because a constant can be added to all player ratings without changing the fit, the sum of all player ratings was set to zero. Since a rating is included for the actual players involved in the set, the set weakness and opposition percentage variables are no longer needed, but the effects can be calculated later by accumulating the ratings of the competing players. This can be used to check the accuracy of the figures obtained using just the club data.

Rating non-elite tennis players

The results give a rating and consequent ranking of all the players in the section, and could be used for interest, to award best player prizes, or to rate the teams for promotion or relegation. Because of possible privacy issues I have not given complete list of final ratings here. However the ratings generally agreed with player perceptions. Within clubs the players were generally ranked in the order in which they played.

The common home advantage of 0.51 was significant at $p = .026$.

The players in our team were ranked 10, 16, 22, 26 and 46 and their ratings are given in column 3 of Table 4, compared with the ratings given in the paper by using the club figures only which are shown in Column 2.

Table 4 Alternative Ratings of North Balwyn team

Player	Least squares rating using club figures	Least squares rating using all data	Least Absolute value rating using all data	Final Exponential smoothing rating
EC	2.5	1.9	2.6	1.7
GM	-0.1	0.1	1.3	-0.3
SC	1.0	1.0	0.4	1.3
GB	0.1	0.5	0.6	-0.4
JZ	-3.1	-1.9	-2.0	-1.9

While this set of ratings has less spread than the original, the two sets have an almost perfect linear relationship, which suggests that the method of calculating ratings using only the data available to the club provides reasonable estimates of players relative abilities.

The difficulty of playing in certain positions can now be assessed by summing the ratings of the actual players who played in those positions. The average sum of the ratings of the two players for each team in each set of a given standard is shown in Table 5.

Table 5 Set difficulty

Easiness	Set	Sum of ratings
3	1&2	1.56
4	1&3	0.98
5	1&4	0.62
5	2&3	0.12
6	2&4	-0.24
7	3&4	-0.82

The difficulty of the set is decreasing as the easiness increases. Clearly 1 & 4 is much harder than 2 and 3, but if we average these two at 0.36, the drop in ratings is virtually a constant 0.6 for each increase in easiness. Thus the assumption of linearity used in the analysis of the club data is reasonable.

The average for each set of the ratings of the players who played in each position can be calculated, and from 1 to 4 they were 1.21, 0.35, -.23, -.59 with an overall average of 0.19. (Note that summing the relevant ratings gives the final column of Table 5). Clearly the better players do play in the higher positions. But the difference between 1 and 2 (0.86) is greater than between 2 and 3 (0.58) which is in turn greater than the difference between 3 and 4 (0.36). Since the overall average of all 52 players who played any sets was chosen to be 0, the average over sets of 0.19 shows that the better players play more often than the weaker players. To be average you need to be better than average: i.e. to have an expectation of being even in games up over the season you need to be better than the average standard of all the players who play at least once.

Using Microsoft Excel, the ratings can be optimized by minimizing the absolute values of the errors, rather than the squares. This may seem preferable to many players, as it would take less notice of larger errors. This does however produce slightly different ratings. Although the correlation between the two sets of 52 ratings is 0.94, the average move in the rankings is 3.7, with one player rising 18 places from 28 to 10. However this player played only one match, so one might expect volatile ratings in this case. In our team's case, our players are now ranked 7, 17, 22, 25 and 45 but the order has changed. The ratings obtained are given in Column 4 of Table 2, and show that GM has jumped to second place, and SC has dropped to fourth.

It is not clear what aspect of a person's game (consistency maybe?) that would be 'rewarded' or 'penalized' by one rating system over the other.

An exponential smoothing method.

It is probably not feasible to expect an association to perform a complicated regression analysis to rate players. However this still has value for testing the reliability of other methods. There are also some difficulties in applying the methods above. The ratings can only be calculated at the end of the season, or at least once a substantial portion of the season has been completed. They result in a single rating, and give no indication of periods of good or bad form, or improvement or decline. There is also a 'black box' feel, where it may be difficult for many players to see how the ratings relate to their results. A simpler system that can be managed on an ongoing basis would be preferable. A method of rating players using exponential smoothing has been shown to be feasible for elite players (Bedford and Clarke, 2000). These methods give a dynamic rating that is updated each week. They are also conceptually simple. A comparison of the ratings of the two pairs of participating players gives an expected set margin. If the players do better than predicted, their rating goes up; worse than predicted and their rating goes down. Thus we have, where α is a smoothing constant between 0 and 1,

$$\begin{aligned} \text{Predicted set margin} &= \text{player rating} + \text{partner rating} - \text{opponent A rating} - \text{opponent B rating} \\ \text{Updated player rating} &= \text{Previous player rating} + \alpha * (\text{actual set margin} - \text{predicted set margin}) \end{aligned}$$

This simple smoothing method was used to rate all 52 Section 2 players using the association results. Three different settings for the starting value of the parameters were tested. In the first case all the initial ratings were set to zero, and a smoothing constant of 0.2 was used. Values around 0.2 have been used by the author for many years with some success in predicting sporting results (Clarke, 1993). Since one would generally expect better players to play in higher positions, in the second case initial ratings were based on the position the player filled when he first played. Finally the value of α was optimized to give the best fit to the predicted set results. As expected each refinement in the method showed an increase in the correlation of the end of the season exponential smoothing ratings and the least squares regression ratings – see Table 6. Given that the regression ratings are a single rating for the whole season's performance, and the smoothing rating is an estimate of the players rating at the end of the season, these indicate the smoothing method is able to give reasonable ratings. The end of season ratings given to our players using the optimized α are shown in the final column of Table 2, and are very similar to the least squares regression ratings shown in Column 3.

Table 6 Comparison of final ratings by exponential smoothing method and regression

Method	Correlation between exponential smoothing and regression ratings
$\alpha=0.2$, initial ratings all zero.	0.76
$\alpha=0.2$, initial ratings based on position when first played	0.81
Optimized $\alpha=0.08$, initial ratings based on position when first played.	0.85

Conclusion

This paper demonstrates that a simple accumulation of a tennis player's scores can produce misleading results. However using a statistical package or just a spreadsheet clubs can overcome a lack of data, and produce ratings that allow for partner, position played and strength of opposition team. Associations can do better, and allow also for individual opponent. Both methods found evidence of a significant home advantage in this local tennis competition. However such methods involve using statistical expertise and packages that may be beyond the resources of many clubs. A rating system using exponential smoothing needs at best an electronic calculator. Such a system would be useful for players to compare abilities, clubs to select players in teams, associations to group teams and have a rational promotion and relegation system. Previous work

has shown it works for elite tennis players participating in the ATP tournaments. This paper has demonstrated its ability to give reasonable results at the other end of the scale, in non-elite doubles competition. However the set of players in both these studies used players of similar or limited range of abilities. The Elo chess rating system also uses exponential smoothing and gives ratings from beginner to world champion. There is no reason why such a system could not apply in tennis from beginners to circuit players. There is a need to further develop and test the ability of this rating method to handle a wider range of player abilities.

References

- Bedford A (2004). Predicting women's World Cup handball match outcomes using optimised ratings models. In: Morton H & Ganesalingam S (eds). *Seventh Australian Conference on Mathematics and Computers in Sport*, Massey University: Palmerston Nth., New Zealand, pp. 66-74.
- Bedford AB and Clarke SR (2000). A comparison of the ATP ratings with a smoothing method for match prediction. In: Cohen G and Langtry T (eds). *Fifth Australian Conference on Mathematics and Computers in Sport*. University of Technology Sydney: Sydney, Australia, pp. 43-51.
- Clarke, S. R. (1993). Computer forecasting of Australian Rules football for a daily newspaper. *J Opl Res Soc.* **44**: 753-759.
- Clarke, SR (1994). An adjustive rating system for tennis and squash players. In: de Mestre N (ed). *Second Conference on Mathematics and Computers in Sport*. Bond University: Gold Coast, Australia, pp. 43-50.
- Elo AE (1978). *The rating of chess players, past and present*. London, Batsford.
- Ladds M and Bedford A (2008). A player rating model in men's badminton. In: Hammond J (ed). *Ninth Australasian Conference on Mathematics and Computers in Sport*. MathSport: Tweed Heads, Australia. pp. 60-67.
- Stefani RT. (1997). Survey of the major world sports rating systems. *J Appl Stat* **24**: 635-646.
- Stefani RT (1998). A taxonomy and survey of sports rating systems. In Bennett J (ed). *Statistics in Sport*. Arnold: New Jersey. pp. 253-258.
- Wright MB (2009). 50 years of OR in sport. *J Opl Res Soc.* In press.

Acknowledgements

A short version of this paper was presented at the IMA Second International Conference on Mathematics in Sport at Groningen, 1990.

Thanks to David Burn and The Eastern Melbourne Veterans Tennis Association for supplying scorecards.