Forecasting Tennis Match Results Using the Bradley-Terry Model

Aisha Fayomi, Rizwana Majeed, Ali Algarni, Sohail Akhtar, Farrukh Jamal, and Jamal Abdul Nasir

1Faculty of Science Department of Statistics, King Abdulaziz University, Jeddah, Saudi Arabia
2Department of Statistics, Government College University, Lahore, Pakistan
3Department of Mathematics and Statistics, University of Haripur, Haripur, Pakistan
4Department of Statistics, The Islamia University of Bahawalpur, Pakistan

Correspondence should be addressed to Farrukh Jamal; farrukh.jamal@iub.edu.pk

Received 17 December 2021; Revised 15 February 2022; Accepted 16 February 2022; Published 8 March 2022

1. Introduction

Forecasting has been playing an important role in different fields of life, i.e., in decision-making activities of management, to predict uncertain events within an organization, in weather forecasting, in flood forecasting, etc. Stakeholders involved in betting market take advantage of tennis forecasting directly or indirectly. Winning probability calculated using forecasting models helps the bettors in deciding whether to place a bet or not. Keeping in view the importance of tennis forecasting, the Bradley-Terry model is used to model men’s professional tennis for predicting match outcomes in tennis matches of men’s singles. Model coefficients are estimated using data from January 2019 to September 2020 of 3439 matches. Ratings for each player are calculated using model coefficients. Player rankings are then calculated using these ratings. Comparison of model rankings with ATP rankings has shown satisfactory results. Winning probability for each player is calculated using model coefficients and ratings. These probability predictions are evaluated against four measures of performance. The results reveal that surface on which a game is played on contributes significantly towards a player’s performance. Due to this impact of the surface, our model has produced superior player rankings for certain players who had been ranked very low in official ATP rankings. According to most of the performance measures, the model has shown good results for clay court data which are closely followed by hard court data. To calculate return on investment, model results are compared with the bookmakers’ average odds and best available odds. It has been found that return on investment for a fitted model is highest in the case of clay court data in comparison to bookmaker’s average odds and best odds.
In this study, paired comparisons for forecasting tennis match results will be done. Several methods are available to forecast match outcomes in tennis. For example, Bradley-Terry-type model was employed to forecast match results for the top tier of men’s professional tennis, the ATP tour [12]. Similarly, a high-dimensional dynamic model was used to forecast tennis match results [13]. Furthermore, tennis match outcomes were forecasted using within-match betting markets [14]. Kovalchik compared predictive performance of different forecasting models for predicting the tennis outcomes [15].

Paired comparison method is a handy decision-making tool, which describes and compares the values to each other. For making decision in the absence of any objective data, the paired comparison method is a very handy tool. This method is also known as the pairwise comparison [16]. In pairwise comparison, the items which we have to compare are judged in pairs to check which one is preferred and to rank the items being compared [17, 18]. The main reason for applying pairwise comparison methods is simplicity in judging two items instead of several items at once. This method can be used in various situations, i.e., in subjective evaluation criteria or when important priorities are unclear, for example, modeling competitive ability in sports and choice behavior, i.e., preference of the democratic presidential candidate to the Republican candidate or the preference of one soft drink to another.

1.1. Rating. A rating is a numerical value assigned to each team or player created on their contributions, while a ranking is the ordinal position based on the ratings [19]. The rating plays an important role in forecasting and prediction. It is used to indicate the strength of items or objects relative to each other and is used in almost every field. The player’s rating is a system used to rate players individually. The rating can be done domestically or internationally, and each of the teams/players is ranked according to their rating or individual performance. The focus of this research is on the player’s ranking to predict match outcomes [20, 21].

1.2. Betting Odds and Betting Strategies. Tennis bets are mainly placed in two kinds of markets: bookmakers or betting exchanges. In the former type, odds are offered by a bookmaker and customers who accept odds and directly place money at these odds against the bookmaker [22]. In betting exchange, odds are offered by customers and are placed against each other; for each paired bet, exchange simply takes a small commission. Generally, odds are more favorable in exchange markets, but due to the limitation of historical data on such odds, the models are mostly compared against traditional bookmaker’s odds in research. The problem in making bets is to find successful bets, where the considered probability of occurrence is greater than the corresponding probability calculated by the bookmakers’ odds, to get a positive expected return. A statistical model capable to predict probabilities of the results of tennis matches accurately can form a profitable betting strategy. Bets can be placed on different events relevant to various aspects of tennis match, both before starting the match and during the match. However, bets placed before starting the match will be considered here [23, 24].

The purpose of betting strategies is to take advantage from those cases in which odds are undervalued. In these cases, the actual probability of the event in an outcome occurring is greater than the implied probability calculated using odds. If the predictive model is available, a simple betting strategy is placing a unit bet when the model shows that the odds are undervalued.

2. Data and Model

For the current study, data on tennis matches of men’s singles have been obtained from tennis-data.co.uk. The dataset consists of 3439 men’s single matches from January 2019 to September 2020. The dataset contains information about tournament name; date of match; location; ATP rankings; name of the winning and losing player; surface (carpet, clay, hard court, or grass); the results in sets, games, and matches; and the series. “Series” relates to the importance of tournament in terms of the prize money available and ranking points; it is divided into ATP250, ATP500, Grand Slam, Masters 1000, and Masters Cup. Results of previous 52 weeks of tournament competition have been used to derive ATP rankings. Awarded points are based on prestige of the tournament and on the progress of player. Records for the odds of bookmakers are also available at the website of tennis data for each game, which are utilized to measure the forecasting performance.

The purpose of the study is to forecast the result of the tennis matches between two competing players. The available data has been analyzed under Bradley-Terry modeling framework using the BradleyTerry2 package in R to obtain ranking of tennis players for the future.

2.1. Model. The Bradley-Terry model is a simple, popular, and widely used method for handling data on paired comparisons which is used to find the probabilities of the potential outcomes when individuals or objects are judged against each other in the form of pairs. This model is used to predict matches and tournaments and for testing the efficiency of betting markets. The default modeling approach to the statistical analysis of tennis matches is also based on the Bradley-Terry model [25, 26].

Suppose a set of $K$ elements are compared with each other in the form of pairs. To find the probability that “$i$” is preferred to “$j$,” for two elements “$i$” and “$j$” taken from this set, the following model was suggested by Bradley and Terry in 1952.

$$P(i \text{ is preferred to } j) = \frac{\lambda_i}{\lambda_i + \lambda_j},$$

where the parameter $\lambda_i > 0$ belongs to element $l \in \{1, 2, \ldots, K\}$ which represents rating and is denoted by $\lambda = \{\lambda_l\}_{l=1}^{K}$.
This model has many applications in different fields as mentioned by Hunter; an extensive bibliography published by Davidson and Farquhar (1976) on the method of paired comparisons also includes a considerable amount of work [27, 28]. For example, it has been adopted to rank players by the European Go Federation and the World Chess Federation and Hastie and Tibshirani suggested that this model is a standard approach for constructing multiclass classifiers using the results of binary classifiers [29]. Different extensions of this model have been suggested to handle draws, home advantage, team comparisons, and multiple comparisons [30, 31]. In particular, the Plackett-Luce model was developed which is the popular extension for multiple comparisons, has been used for ranking of multiple objects, defined a prior distribution of permutations, and has also been used for choice models [32]. In the monographs of David [33] and Diaconis [34], detailed discussion about the statistical foundations of these models has also been provided.

### 3. Ranking of Tennis Players on All Surfaces

Keeping in view the scope of the current study, it is of great interest to utilize the abilities of players $\{a_i\}$ to model a new ranking system as an alternative to official rankings published by ATP and compare both the ranking systems. Model parameters have been estimated using data for the period from January 2019 to September 2020, to obtain ranking of tennis players, assigning equal weights to all matches played on any surface.

Rankings produced by the model at the end of September 2020 for the top fifteen players have been shown in Table 1 along with ATP rankings. The rankings produced by the model and the ATP rankings are more or less the same for the top 7 players, except a few discrepancies such as Thiem D. and Federer R., whose ATP ranking is 3 and 4, respectively, whereas ranking of these players according to our model is 4 and 3, respectively. This change in ranking of Thiem D. and Federer R. is due to the difference in the performance of these players on different surfaces; i.e., Thiem D. performs well on clay whereas clay is Federer’s least favored surface and he has managed to win only one title at the Roland Garros (French Open). Moreover, Federer R. has played on all surfaces, i.e., hard court, clay court, and grass court whereas Thiem D. never won a game on grass court in our dataset. Thiem D. has won more games, i.e., 63 in our dataset, than Federer R., who has managed to win 58 games. Federer R. has won less games than Thiem D., but he has played on all surfaces that contributed in his better overall performance than Thiem D. It is also a historical fact that Federer R. has won more Wimbledon (grass court game) titles than any other player. Therefore, it is quite evident that grass is his most preferred playing surface.

Similarly, Berrettini M. has been ranked 7 in our model whereas he is at number 8 according to official ATP rankings. Zverev A., who is at number 7 in official ATP ranking, has been ranked 19 in our model. This is due to the fact that Zverev A. has managed to win only two games on grass court. Although Zverev A. has won more games, i.e., 53 in total than Berrettini M., who managed to win 49 games, Berrettini M.

### Table 1: Model ranking and ATP ranking of top 20 players on all surfaces.

| Model rank | Players          | Estimates | Rating | Std. error | ATP ranking |
|------------|------------------|-----------|--------|------------|-------------|
| 1          | Djokovic N.      | 0.2725    | 1.3132 | 0.5181     | 1           |
| 2          | Nadal R.         | 0.0000    | 1.0000 | 0.0000     | 2           |
| 3          | Federer R.       | -0.1672   | 0.8460 | 0.5154     | 4           |
| 4          | Thiem D.         | -0.231    | 0.7937 | 0.4977     | 3           |
| 5          | Medvedev D.      | -0.9094   | 0.4028 | 0.4928     | 5           |
| 6          | Tsitsipas S.     | -1.0045   | 0.3662 | 0.4585     | 6           |
| 7          | Berrettini M.    | -1.3522   | 0.2587 | 0.5062     | 8           |
| 8          | Rublev A.        | -1.3591   | 0.2569 | 0.4869     | 12          |
| 9          | De Minaur A.     | -1.4134   | 0.2433 | 0.5524     | 28          |
| 10         | Raonic M.        | -1.4193   | 0.2419 | 0.5577     | 21          |
| 11         | Wawrinka S.      | -1.4465   | 0.2354 | 0.5167     | 17          |
| 12         | Schwartzman D.   | -1.474    | 0.2290 | 0.503      | 14          |
| 13         | Struff J. L.     | -1.5372   | 0.2150 | 0.4939     | 32          |
| 14         | Monfils G.       | -1.5478   | 0.2127 | 0.5247     | 9           |
| 15         | Opelka R.        | -1.5483   | 0.2126 | 0.5134     | 36          |
| 16         | Bautista Agut R. | -1.5949   | 0.2029 | 0.495      | 10          |
| 17         | Kyrgios N.       | -1.6348   | 0.1950 | 0.5648     | 42          |
| 18         | Krajnovic F.     | -1.7045   | 0.1819 | 0.5476     | 29          |
| 19         | Zverev A.        | -1.7391   | 0.1757 | 0.5187     | 7           |
| 20         | Dimitrov G.      | -1.7691   | 0.1705 | 0.5243     | 20          |

3.1. Winning Probabilities on All Surfaces. Using values of estimates from model fit given in Table 2 and Figure 3, without considering the surface game played on, at the end of 2020, the estimated winning probabilities of competing players can be calculated using the following equation:

$$\hat{p}_{ij} = \frac{\exp (\hat{a}_i - \hat{a}_j)}{1 + \exp (\hat{a}_i - \hat{a}_j)}.$$  

(2)
**Figure 1:** Comparison of model ranking with ATP ranking for top 50 tennis players on all surfaces.

**Figure 2:** Estimated relative abilities of tennis players on all surfaces.

**Table 2:** Winning probabilities of players on all surfaces. DN: Djokovic N.; NR: Nadal R.; FR: Federer R.; TD: Thiem D.; MD: Medvedev D.; TS: Tsitsipas S.; ZA: Zverev A.; BM: Berrettini M.; DMA: De Minaur A.; RM: Raonic M.

|       | DN  | NR   | FR   | TD   | MD   | TS   | BM   | RA   | DMA  | RM   |
|-------|-----|------|------|------|------|------|------|------|------|------|
| DN    | 0.568| 0.608| 0.623| 0.765| 0.782| 0.835| 0.836| 0.844| 0.844|      |
| NR    | 0.432| 0.542| 0.557| 0.713| 0.732| 0.794| 0.796| 0.804| 0.805|      |
| FR    | 0.392| 0.458| 0.516| 0.677| 0.698| 0.766| 0.767| 0.777| 0.778|      |
| TD    | 0.377| 0.443| 0.484| 0.663| 0.684| 0.754| 0.755| 0.765| 0.766|      |
| MD    | 0.235| 0.287| 0.323| 0.337| 0.524| 0.609| 0.611| 0.623| 0.625|      |
| TS    | 0.218| 0.268| 0.302| 0.316| 0.476| 0.586| 0.588| 0.601| 0.602|      |
| BM    | 0.165| 0.206| 0.234| 0.246| 0.391| 0.414| 0.502| 0.515| 0.517|      |
| RA    | 0.164| 0.204| 0.233| 0.245| 0.389| 0.412| 0.498| 0.514| 0.515|      |
| DMA   | 0.156| 0.196| 0.223| 0.235| 0.377| 0.399| 0.485| 0.486| 0.501|      |
| RM    | 0.156| 0.195| 0.222| 0.234| 0.375| 0.398| 0.483| 0.485| 0.499|      |

Winning probabilities of players in the left-hand column over the players in the top row, on all surfaces.
To check the statistical significance of the difference between Djokovic N. and Nadal R., difference of the estimates \((\alpha_1 - \alpha_2) = 0.2725\) is compared with its standard error \((SE = 0.5181)\) having \(p\) value \(= 0.5990\) indicating that the difference between the performance of Djokovic N. and Nadal R. is insignificant.

A 95% confidence interval for \(\alpha_1 - \alpha_2\) is

\[
0.2725 \pm 1.96(0.5181) \text{ or } (-0.743, 1.288),
\]

and the 95% confidence interval for probability \(\pi_{12}\) of Djokovic to win obtained using the above interval is \((0.32, 0.78)\).

### 4. Ranking of Tennis Players on Hard Court

In our model ranking for all surfaces, we saw some discrepancies which were actually due to the difference of performance on different surfaces. Some players perform well on hard court, some perform well on clay court, and some give better results on grass court. Keeping in view this fact, a separate analysis has been carried out for hard court and clay court competitions. This separate analysis has proved that surface has a very strong impact on a player's performance. Especially, we found that true ability of a player is judged by his performance on hard court. In the following, results from our model for hard court are presented.

Ranking produced by the model at the end of September 2020 for the top fifteen players on hard court has been shown in Table 3 along with ATP rankings (Figure 4). The model ranking and ATP ranking for the 1st six players are the same which shows the model has produced good results when surface effect has been taken into consideration. Raonic M. whose ranking is 21 according to ATP ranking has been ranked 7 by the model. This is due to the reason that Raonic M. has played and won more matches on hard court than any other surface. On clay court, he has played and won only 1 match, and on grass court, he managed to win 8 matches. Similarly, De Minaur A. has won 35 matches on hard court out of a total of 38 he has won. He has won

Table 3: Model ranking and ATP ranking of top 20 players on hard court.

| Model rank | Player         | Estimate | Rating | Std. error | ATP ranking |
|------------|----------------|----------|--------|------------|-------------|
| 1          | Djokovic N.    | 0.1416   | 1.1521 | 0.6918     | 1           |
| 2          | Nadal R.       | 0.0000   | 1.0000 | 0.0000     | 2           |
| 3          | Thiem D.       | -0.0556  | 0.9459 | 0.6985     | 3           |
| 4          | Federer R.     | -0.2224  | 0.8006 | 0.7063     | 4           |
| 5          | Medvedev D.    | -0.6465  | 0.5239 | 0.6387     | 5           |
| 6          | Tsitsipas S.   | -1.0456  | 0.3515 | 0.6267     | 6           |
| 7          | Raonic M.      | -1.0508  | 0.3497 | 0.7172     | 21          |
| 8          | De Minaur A.   | -1.2714  | 0.2804 | 0.6871     | 28          |
| 9          | Opelka R.      | -1.3611  | 0.2564 | 0.6866     | 36          |
| 10         | Kyrgios N.     | -1.3994  | 0.2467 | 0.7363     | 42          |
| 11         | Struff J. L.   | -1.4251  | 0.2405 | 0.6964     | 32          |
| 12         | Monfils G.     | -1.4462  | 0.2355 | 0.6782     | 9           |
| 13         | Berrettini M.  | -1.4814  | 0.2273 | 0.7028     | 8           |
| 14         | Rublev A.      | -1.4879  | 0.2259 | 0.6456     | 12          |
| 15         | Wawrinka S.    | -1.5668  | 0.2087 | 0.6779     | 17          |
| 16         | Bautista Agut R.| -1.6233  | 0.1973 | 0.6697     | 10          |
| 17         | Dimitrov G.    | -1.6310  | 0.1957 | 0.6989     | 20          |
| 18         | Krajnovic F.   | -1.6960  | 0.1834 | 0.6986     | 29          |
| 19         | Carreno Busta P.| -1.7105  | 0.1808 | 0.6790     | 18          |
| 20         | Zverev A.      | -1.7151  | 0.1800 | 0.7016     | 7           |

The estimated winning probability of Djokovic N. vs. Nadal R. is

\[
\hat{\pi}_{12} = \frac{\exp (\hat{\alpha}_1 - \hat{\alpha}_2)}{1 + \exp (\hat{\alpha}_1 - \hat{\alpha}_2)} = \frac{\exp (0.2725 - 0)}{1 + \exp (0.2725 - 0)} = 0.568. \tag{3}
\]
only 2 matches on grass court and 1 match on clay court. Because of this reason, he has been ranked 8 in our model. Likewise, Opelka R. whose ATP rank is 36 has been ranked 9 in our model due to the reason that he has played and won more matches on hard court than any other surface. Opelka R. has won 32 matches on hard court out of a total of 36 wins. He has won only two matches on clay court and two on grass court.

The ability of players can also be estimated using comparison intervals calculated using quasistandard errors as depicted in Figure 5. This method not only provides comparison with the reference player, but readily made comparisons between any pair of competing players are also provided.

4.1. Winning Probabilities on Hard Court. Using values of estimates from model fit given in Table 4 and Figure 6, estimated winning probabilities of competing players can be calculated using the following equation:

\[ \pi_{ij} = \frac{\exp(\hat{\alpha}_i - \hat{\alpha}_j)}{1 + \exp(\hat{\alpha}_i - \hat{\alpha}_j)} \]  

Figure 4: Comparison of model ranking with ATP ranking for top 50 tennis players on hard court.

Figure 5: Estimated relative abilities of tennis players on hard court only.

| Players | Model ranking | ATP ranking |
|-----------------|-----------------|-----------------|

Table 4: Winning probabilities of players on hard court. DN: Djokovic N.; NR: Nadal R.; TD: Thiem D.; FR: Federer R.; MD: Medvedev D.; TS: Tsitsipas S.; RM: Raonic M.; DMA: De Minaur A.; OR: Opelka R.; KN: Kyrgios N.
The estimated winning probability of Djokovic N. vs. Nadal R. is

\[ \pi_{12} = \frac{\exp (\hat{\alpha}_1 - \hat{\alpha}_2)}{1 + \exp (\hat{\alpha}_1 - \hat{\alpha}_2)} = \frac{\exp (0.1416 - 0)}{1 + \exp (0.1416 - 0)} = 0.535. \]  

To check the statistical significance of the difference between Djokovic N. and Nadal R., difference of the estimates \((\hat{\alpha}_1 - \hat{\alpha}_2) = 0.1416\) is compared with its standard error \((SE = 0.6918)\) having \(p\) value = 0.8378 indicating that the difference between performance of Djokovic N. and Nadal R. is insignificant.

A 95% confidence interval for \(\alpha_1 - \alpha_2\) is

\[ 0.1416 \pm 1.96(0.6918) \text{ or } (-1.2143, 1.4975), \]  

and the 95% confidence interval for probability \(\pi_{12}\) of Djokovic to win obtained using the above interval is \((0.23, 0.82)\).

### 5. Ranking of Tennis Players on Clay Court

Rankings produced by the model at the end of September 2020 for the top fifteen players on clay court have been shown in Table 5 along with ATP rankings (Figure 7). The model ranking and ATP ranking for first four players are the same which shows the model has produced good results when surface effect has been taken into consideration. It is pertinent to note that Wawrinka S., who was ranked 11 and 15 in all surface model and hard-court model, respectively, is now ranked 5 in the clay court model, whereas Garin C., who was ranked 26 and 42 in all surface models and hard-court model, respectively, is now ranked 6 in the clay court model. This significant upward change in ranking is due to the fact that Wawrinka S. and Garin C. have performed well on clay court as compared to hard court. Wawrinka S. has won 8 matches out of 14 on clay court while Garin C. has won 32 matches out of 43 on clay court.

The ability of players can also be estimated using comparison intervals calculated using quasistandard errors as depicted in Figure 8. This method not only provides comparison with the reference player, but readily made comparisons between any pair of competing players are also provided.

#### 5.1. Winning Probabilities on Clay Court

Using values of estimates from model fit given in Table 6 and Figure 9, estimated winning probabilities of competing players on clay court.
Table 6: Winning probabilities of players on clay court. DN: Djokovic N.; NR: Nadal R.; TD: Thiem D.; FR: Federer R.; WS: Wawrinka S.; GC: Garin C.; RA: Rublev A.; TS: Tsitsipas S.; RC: Ruud C.; SD: Schwartzman D.

|     | DN | NR  | TD  | FR  | WS  | GC  | RA  | TS  | RC  | SD  |
|-----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| DN  | 0.528 | 0.607 | 0.682 | 0.713 | 0.731 | 0.737 | 0.739 | 0.751 | 0.775 |
| NR  | 0.472 | 0.580 | 0.657 | 0.690 | 0.708 | 0.715 | 0.717 | 0.730 | 0.755 |
| TD  | 0.393 | 0.420 | 0.582 | 0.617 | 0.638 | 0.645 | 0.647 | 0.662 | 0.691 |
| FR  | 0.318 | 0.343 | 0.418 | 0.537 | 0.559 | 0.567 | 0.569 | 0.585 | 0.616 |
| WS  | 0.287 | 0.310 | 0.383 | 0.463 | 0.522 | 0.530 | 0.533 | 0.549 | 0.581 |
| GC  | 0.269 | 0.292 | 0.362 | 0.441 | 0.478 | 0.509 | 0.511 | 0.527 | 0.559 |
| RA  | 0.263 | 0.285 | 0.355 | 0.433 | 0.470 | 0.491 | 0.502 | 0.518 | 0.551 |
| TS  | 0.261 | 0.283 | 0.353 | 0.431 | 0.467 | 0.489 | 0.498 | 0.516 | 0.549 |
| RC  | 0.249 | 0.270 | 0.338 | 0.415 | 0.451 | 0.473 | 0.482 | 0.484 | 0.533 |
| SD  | 0.225 | 0.245 | 0.309 | 0.384 | 0.419 | 0.441 | 0.449 | 0.451 | 0.467 |

Winning probabilities of players in the left-hand column over the players in the top row, on clay court.
Figure 9: Match by match comparison of predicted probability with implied probability from bookmakers’ average and best odds on clay court.

Table 7: Measures of predictive performance for fitted model.

| Measures of performance | All surfaces | Fitted model | Clay court |
|-------------------------|-------------|--------------|------------|
| Classification accuracy $m_1$ | 66.49% | 67.89% | 73.27% |
| Average probability $m_2$ | 0.5871 | 0.6134 | 0.6499 |
| Average log probability $m_3$ | -0.5993 | -0.5744 | -0.5142 |
| Return on investment $m_4$ | -8.60% | 8.20% | 19.80% (Avg. odds) |
|                          | 8.81% | 21.82% | 22.77% (Max. odds) |

Figure 10: Measures of predictive performance for the fitted model.
The estimated winning probability of Djokovic N. vs. Nadal R. is

\[
\tilde{\pi}_{12} = \frac{\exp (\hat{\alpha}_1 - \hat{\alpha}_2)}{1 + \exp (\hat{\alpha}_1 - \hat{\alpha}_2)} = \frac{\exp (0.1119 - 0)}{1 + \exp (0.1119 - 0)} = 0.528. 
\]

To check the statistical significance of the difference between Djokovic N. and Nadal R., difference of the estimates \((\hat{\alpha}_1 - \hat{\alpha}_2 = 0.1119)\) is compared with its standard error \((SE = 0.8803)\) having \(p\) value = 0.8988 indicating that the difference between performance of Djokovic N. and Nadal R. is insignificant.

A 95% confidence interval for \(\alpha_1 - \alpha_2\) is

\[
0.1119 \pm 1.96(0.8803) \text{ or } (-1.6135, 1.8373),
\]

and the 95% confidence interval for probability \(\pi_{12}\) of Djokovic to win obtained using the above interval is \((0.17, 0.86)\).

### 6. Measures of Performance

In Table 7, four measures of predictive performance are presented. According to all performance measures, the model is overall well fitted for all data, but for clay court matches, it shows best results for which model accuracy is 73.27%, higher than any other model in our data. Moreover, average probability and average log probability which are two important performance measuring criteria also show that the model is well fitted for clay court data. Similarly, return on investment is highest for clay court matches on average bookmakers’ average odds. From Table 7, it is obvious that when the effect of surface is considered, the model provides good return on investment as model predictions have been compared against bookmaker’s average odds and best available odds.

From Figure 10, it is quite clear that for all the measures of predictive performance, the model is overall well fitted. It shows best results for clay court in the fitted model, closely followed by results for hard court. So, it can be concluded that surface has significant impact on the performance of the players and ultimately on the outcome of the match.

### 7. Closing Remarks

The objective of this study was to introduce a model for predicting match outcomes in men’s professional tennis using ratings obtained from the fitted model to rank players keeping in view the effect of different surfaces. In studies conducted previously on this topic, the modeling approach was restricted to the models in which information on official rankings was used. In the current study, the Bradley-Terry approach has been applied on historical data on men’s singles results to obtain match forecasts which are more accurate according to several criteria because the evidence from soccer, tennis, and golf suggests that the official rankings of teams and players although useful predictors do not contain the entire information required for forecasting results. To achieve the objectives of the study and implement the Bradley-Terry approach, historical data on men’s singles was obtained from January 2019 to September 2020 of 3439 matches. The dataset was categorized into three major categories with respect to surface type: all surfaces (consisting of matches’ data on hard court, clay court, and grass court), hard court, and clay court. The Bradley-Terry model was applied on the dataset for each surface category. In our analysis on dataset of all surface types, equal weightage has been given to hard court, clay court, and grass court. Analysis for hard court and clay court has been done separately for each surface.

The analysis revealed that surface of the match has significant impact on the performance of the players. Due to this impact of the surface, our model produced superior player rankings for certain players who had been ranked very low in official ATP rankings. Using coefficients of model, ratings of players were found for each category of surface. These ratings were then used to calculate winning probabilities of players. To check the adequacy of the model,
four performance criteria were used to measure the performance of the model predictions. According to most of the performance measures, the model has shown good results for clay court data which are closely followed by hard court data. To calculate return on investment, model results were compared with the bookmakers’ average odds and best available odds. It was found that return on investment was highest for clay court data when compared with both bookmaker’s average odds and best odds.

**Data Availability**

The data used in the article is given therein.

**Conflicts of Interest**

No conflict of interest was declared by the authors.

**Acknowledgments**

The Deanship of Scientific Research (DSR), King Abdul-Aziz University, Jeddah, supported this work, under grant no. G: 358. The authors, therefore, gratefully acknowledge the DSR technical and financial support.

**References**

[1] B. L. Boulier and H. O. Stekler, "Are sports seedings good predictors? an evaluation," *International Journal of Forecasting*, vol. 15, no. 1, pp. 83–91, 1999.

[2] P. K. Newton and J. B. Keller, "Probability of winning at tennis I. Theory and data," *Studies in Applied Mathematics*, vol. 114, no. 3, pp. 241–269, 2005.

[3] W. J. Knottenbelt, D. Spanias, and A. M. Madurska, "A common-opponent stochastic model for predicting the outcome of professional tennis matches," *Computers & Mathematics with Applications*, vol. 64, no. 12, pp. 3820–3827, 2012.

[4] S. Akhtar, P. Scarf, and Z. Rasool, "Rating players in test match cricket," *Journal of the Operational Research Society*, vol. 66, no. 4, pp. 684–695, 2015.

[5] P. Scarf, X. Shi, and S. Akhtar, "Modelling batting strategy in test cricket," in *In Progress in Industrial Mathematics at ECMI 2008*, pp. 481–489, Springer, Berlin, Heidelberg, 2010.

[6] F. J. Klaassen and J. R. Magnus, "Forecasting the winner of a tennis match," *European Journal of Operational Research*, vol. 148, no. 2, pp. 257–267, 2003.

[7] M. De Araujo Fernandes, "Using soft computing techniques for prediction of winners in tennis matches," *Machine Learning Research*, vol. 2, no. 3, pp. 86–98, 2017.

[8] P. Scarf and S. Akhtar, "An analysis of strategy in the first three innings in test cricket: declaration and the follow-on," *Journal of the Operational Research Society*, vol. 62, no. 11, pp. 1931–1940, 2011.

[9] F. Lisi and G. Zanella, "Tennis betting: can statistics beat bookmakers?," *Electronic Journal of Applied Statistical Analysis*, vol. 10, no. 3, 2017.

[10] S. Akhtar and P. Scarf, "Forecasting test cricket match outcomes in play," *International Journal of Forecasting*, vol. 28, no. 3, pp. 632–643, 2012.

[11] P. Scarf, X. Shi, and S. Akhtar, "On the distribution of runs scored and bat strategy in test cricket," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 174, no. 2, pp. 471–497, 2011.

[12] I. McHale and A. Morton, "A Bradley-Terry type model for forecasting tennis match results," *International Journal of Forecasting*, vol. 27, no. 2, pp. 619–630, 2011.

[13] S. J. Koopman and R. Lit, "The analysis and forecasting of tennis matches by using a high dimensional dynamic model," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 182, no. 4, pp. 1393–1409, 2019.

[14] S. Easton and K. Uylangco, "Forecasting outcomes in tennis matches using within-match betting markets," *International Journal of Forecasting*, vol. 26, no. 3, pp. 564–575, 2010.

[15] S. A. Kovalchik, "Searching for the GOAT of tennis win prediction," *Journal of Quantitative Analysis in Sports*, vol. 12, no. 3, pp. 127–138, 2016.

[16] R. A. Bradley, "14 paired comparisons: some basic procedures and examples," *Handbook of Statistics*, vol. 4, pp. 299–326, 1984.

[17] N. Dingle, W. Knottenbelt, and D. Spanias, "On the (page) ranking of professional tennis players," in *Computer Performance Engineering*, pp. 237–247, Springer, Berlin, Heidelberg, 2012.

[18] G. H. Kuper, G. Sierksma, and F. C. Spiekman, *Using tennis rankings to predict performance in upcoming tournaments* (No. 14034-EEF), University of Groningen, Research Institute SOM (Systems, Organisations and Management), 2014.

[19] S. R. Clarke and D. Dyte, "Using official ratings to simulate major tennis tournaments," *International Transactions in Operational Research*, vol. 7, no. 6, pp. 585–594, 2000.

[20] L. Knorr-Held, "Dynamic rating of sports teams," *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 49, no. 2, pp. 261–276, 2000.

[21] Š. Lyócsa and T. Výrost, "To bet or not to bet: a reality check for tennis betting market efficiency," *Applied Economics*, vol. 50, no. 20, pp. 2251–2272, 2018.

[22] M. J. Dixon and S. G. Coles, *Modelling association football scores and inefficiencies in the football betting market*, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 46, no. 2, pp. 265–280, 1997.

[23] T. Barnett and S. R. Clarke, "Combining player statistics to predict outcomes of tennis matches," *IMA Journal of Management Mathematics*, vol. 16, no. 2, pp. 113–120, 2005.

[24] N. Dewant and J. Gillard, "Corrigendum to: Using Bradley-Terry models to analyse test match cricket," *IMA Journal of Management Mathematics*, vol. 30, no. 2, 2019.

[25] D. Firth, "Bradley-Terry models inR," *Journal of Statistical Software*, vol. 12, no. 1, pp. 1–12, 2005.

[26] D. R. Hunter, "MM algorithms for generalized Bradley-Terry models," *The Annals of Statistics*, vol. 32, no. 1, pp. 384–406, 2004.

[27] R. R. Davidson and P. H. Farquhar, "A bibliography on the method of paired comparisons," *Biometrics*, pp. 241–252, 1976.

[28] T. Hastie and R. Tibshirani, *Classification by pairwise coupling*, In Advances in neural information processing systems, 1998.

[29] P. V. Rao and L. L. Kupper, "Ties in paired-comparison experiments: a generalization of the Bradley-Terry model," *Journal of the American Statistical Association*, vol. 62, no. 317, pp. 194–204, 1967.

[30] A. Agresti and M. Kateri, *Categorical Data Analysis*, Springer, 2011.
[31] R. L. Plackett, “The analysis of permutations,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 24, no. 2, pp. 193–202, 1975.

[32] P. Diaconis, “Group representations in probability and statistics,” *Lecture notes-monograph series*, vol. 11, 1988.

[33] H. A. David, “Note on “an extreme null distribution approach to the problem of paired comparisons” by Particia S. Costello and Douclas A. Wolfe,” *Communications in Statistics-Theory and Methods*, vol. 17, no. 11, pp. 4005–4009, 1988.

[34] P. Diaconis, “Group representations in probability and statistics,” *Lecture notes-monograph series*, vol. 11, pp. 1–192, 1988.