Kevin Haugh
Scott Powers
Stats 50: Mathematics of Sports
6/2/16

Analysis of Clutch Performance in Men's Professional Tennis

INTRODUCTION

For my project, I analyzed tennis data to determine which players have the best "clutch" performance. In most sports, not only does it matter how many points a player scores, but when the points are scored is also very important. This is especially true in tennis, as you can win more total points than your opponent, yet still lose the match if you are not winning the important ones; if you can't win on match point, you will never win a match. For this project, I have defined clutch performance in tennis as "a player's ability to win the important points."

In sports, players that are "clutch" are those who perform well when the pressure is on, and are fairly easy to pick out just by watching, but it is a much harder task to rank players objectively by their clutch performance. Though statistical analysis is not often used to analyze tennis matches, tennis actually provides a fairly nice platform for statistical analysis because it does not involve interactions with teammates, states can be clearly defined by the current set, game, and point, and each state starts out the same way: with a server and a returner on the opposite sides of the court.

Despite this, I have found there are not many academic studies on clutch performance in tennis. The ones that do exist generally attempt to evaluate clutch performance based on game point and break point win percentage alone, such as a study by Glen Hill of the Tactical Tennis blog (Hill 2013). Both game and break point are important, as they are opportunities to win games, and Hill's study asserts that players with high game point and break point win percentages perform better in clutch situations. However, for my project, I was hoping to go deeper than this. While game and break point are definitely important points within each match, not all game and break points are equally important, and regular points can be even more important than break or game point, depending on the score of the match.

Another study by company Tennis Expert goes further and calculates the theoretical point importance by using simple probability theory ("All Points" 2006). However, tennis players and athletes in general do not always behave in the way a theoretical model might predict because of many factors like pressure, momentum, and mental strength.

METHODS

To evaluate a player's ability to win the important points, there are two steps that need to be taken. First, each point must be assigned a value corresponding to its importance. Second, players are evaluated on their abilities to win points with high importance ratings.

The data I needed to complete these two steps were point-by-point data. I found a relatively solid dataset for public use on GitHub which contains every point from a majority of all Grand Slam matches from 2011-2016 (Sackmann 2016). For each point, the score, the winner of the point, and the server were recorded, along with a great deal of other information regarding what happened in the point (if a player hit a winner, or hit a double fault, etc). The data contain more than 400,000 points, which offered plenty of room for solid statistical analysis. I decided to use Grand Slam data because they are the most important four tournaments of each year and provide matches at roughly the same level of importance.

To investigate my initial research goal, I defined point importance as the increase in probability of winning the match that the player can earn by winning the current point. In other words, at each score:

Score Importance = P(Win match | Win point) – P(Win match)

Initially, calculating the probability that the player wins the match at each score seemed relatively straightforward. I first had to break up each match into states, where the states consisted of every possible scoring combination. In men's Grand Slam matches, matches are won if the player wins 3 sets out of 5. Each set is won by winning 6 games, win by two, with a tiebreak at 6-6. Each game is won by winning 4 points, win by two. However, breaking each point up into one state proved problematic, as it turns out there are a possible 6,669 possible states in a men's Grand Slam match. While my dataset contained over 400,000 point entries, some points are much rarer than others, meaning that several states ended up only having one occurrence in the dataset.

To get around this problem, I instead broke each point up into three states: the set state, the game state, and the point state. Using the formula above, I changed my calculation for score importance to:

SI = (Point Importance)P(Win set | Win game)P(Win match | Win set),

with point importance being the increase in probability a game is won given the point is won. Once I calculated the score importance at each score, I could then evaluate players based on their performance at each score. To do so, I used a normal, regularized Rasch model to calculate beta and delta values for clutch performance in servers and returners, respectively.

RESULTS

First, I'll look at how effective my model was at ranking players according to their respective serve and return clutch performance values. It is difficult to determine the effectiveness of this model in an objective manner, because clutch performance is not a statistic that is widely used, and more importantly, it is not a statistic that has definite, known values like wins or point differential do.

Top 10 Clutch Servers

Rank	Player	Beta
1	Novak Djokovic	1.77%
2	Rafael Nadal	1.41%
3	Stan Wawrinka	1.27%
4	Roger Federer	1.26%
5	Jo-Wilfried Tsonga	1.18%
6	Andy Murray	1.05%
7	Tomas Berdych	0.97%
8	Milos Raonic	0.90%
9	Juan Martin Del Potro	0.87%
10	Andy Roddick	0.74%

Top 10 Clutch Returners

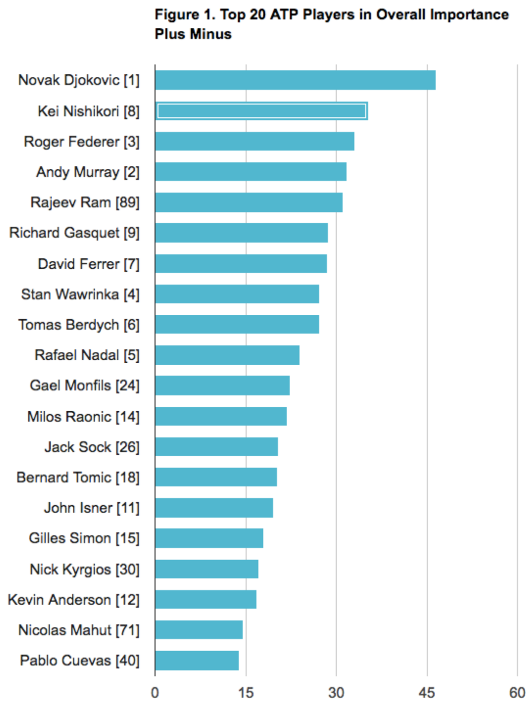
Rank	Player	Delta
1	Novak Djokovic	1.82%
2	Rafael Nadal	1.51%
3	Andy Murray	1.29%
4	Roger Federer	1.21%
5	David Ferrer	1.11%
6	Jo-Wilfried Tsonga	1.02%
7	Tomas Berdych	0.98%
8	Stanislas Wawrinka	0.92%
9	Mardy Fish	0.91%
10	Juan Martin Del Potro	0.76%

Figure 1. Ranking of players by serve and return clutch performance

The beta and delta values for each player represent the boost in match winning percentage that the player wins on an average point above an averagely clutch player on serve and on return. For example, when Novak Djokovic is serving, an average point will result in a 1.77% higher probability of winning the match because he tends to win important points that earn him higher increases in win percentage than the average player. It is a hard metric to try to put into perspective, but it is useful as a metric for comparison.

Looking at these rankings, the first thing that pops out is that Novak Djokovic was far and away the most clutch player in the Grand Slams through the 2011-2016 Grand Slam tournaments. This seems about right, as Djokovic is known for his clutch performance and has been the dominant player for the past several years. In terms of clutch servers, we see Andy Roddick and Milos Raonic, two of the biggest servers in the men's game, high in the ranks of serve clutch performance compared to their return rankings. We also see just the opposite for David Ferrer and Andy Murray, two players lauded for their return games, with their return rankings higher than their serve rankings. Finally, we see Roger Federer at number four in both serve and return clutch performance. While Federer is often considered to be the number one tennis player of all time, his one biggest criticism is converting on important points compared to other greats, which is consistent with his low number four ranking. While none of these criteria can rigorously prove the effectiveness of my model, they at least show that my model is on the right track.

Another way I can validate my model is to compare my results with other available clutch performance rankings. One such ranking model comes from Carl Morris, Emeritus Professor of Statistics at Harvard University. In the most academically rigorous study that I came across, Morris uses a different methodology and trained his model on the 2015 season alone.

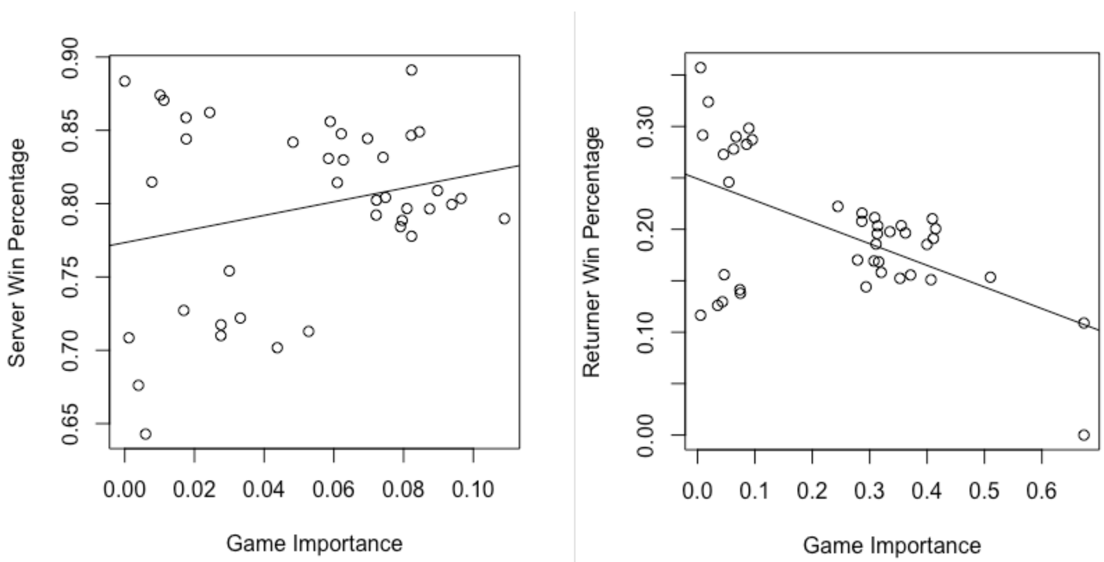


Top 10 Overall Clutch Performers

Rank	Player	Beta
1	Novak Djokovic	1.79%
2	Rafael Nadal	1.46%
3	Roger Federer	1.23%
4	Andy Murray	1.17%
5	Jo-Wilfried Tsonga	1.08%
6	Stan Wawrinka	1.01%
7	Tomas Berdych	0.97%
8	David Ferrer	0.91%
9	Juan Martin Del Potro	0.78%
10	Mardy Fish	0.73%

Figures 2 and 3: Morris' ranking model and my raking model

My results are fairly similar to those of Morris, with mostly the same top 10 players. Two notable exceptions are Kei Nishikori and Rajeev Ram, neither of which made it in the top 10 in my rankings. This is because Ram was not in my dataset, and likely because Nishikori had a breakthrough year in his 2015 season.



Figures 4 and 5: Game importance vs. win percentage of servers and returners

In analyzing my results, I also decided to look at clutch performance among servers and returners in general. In order to see how servers and returners fared on more important points, I plotted the relationships between the game importance (how important the current game is to winning the current set) and the win percentage at each game state. Based on the figures 4 and 5 above, there is a fairly strong negative correlation between game importance and returner win percentage and a positive correlation between game importance and server win percentage. As a whole, servers in professional tennis tend to win points more often than returners because they have the advantage of starting the point out with a strong serve. These results suggest that servers also perform better than returners in clutch situations as well, winning a higher percentage of games as games get more and more important. This result could occur for a variety of reasons. Since players generally server 100-200 serves in a match, servers often don't go all out on serves. However, these results could be occurring because servers are able to recognize when a point is more important and hit a bigger serve.

LIMITATIONS AND EXTENSIONS

Like any statistical model, my model is not without limitations. The first and most noteworthy limitation is that it tends to favor players that get themselves into close matches often. Since point importance gets higher as the score gets closer, players that get themselves into close matches will have more chances to win points with higher importance than players that routinely blow out their opponents. This is evened out somewhat by the fact that players who often are in close matches will have

One important factor that my model did not take into account was the round of the match. As a tournament progresses, matches undoubtedly get progressively more important. For example, a second round match is more important than a first round match, and a final is more important than a semi-final, because winning later matches gets a player closer to winning the championship. Since my dataset did not include data on which round each match was played, I wasn't able to account for this extra variable, but it would be interesting to see how the rankings would change, favoring players that tend to do better in the more important later matches of tournaments.

While ultimately it is most important in terms of clutch performance for a player to win the important points, clutch performance also involves how a player wins a point. A player can win a point simply by luckily having the opponent double fault, but having the opponent double fault says nothing about the player. On the other hand, hitting a winner is much more clutch, because it is something the player had to force to happen. How a player wins a point could potentially be another factor taken into account to determine clutch performance, though the challenge would be to quantify the level of clutch performance of each event.

Works Cited

"All Points Are Not Equal." Tennis Expert Software. 27 June 2006. Web. <http://tennis-expert.narod.ru/st14.htm>

Hill, Glen. "Do Clutch Players Exist? - Tactical Tennis." Tactical Tennis. 18 Jan. 2013. Web. <http://www.tacticaltennisblog.com/clutch-players/>.

Morris, Carl. "Quantifying Clutch Performance." On the T. 27 Dec. 2015. Web. <http://on-the-t.com/2015/12/27/Quantifying-Clutch-Performance/>.

Sackmann, J. (2016). “Point-by-point data for Grand Slams, 2011-15” [Software]. <https://github.com/JeffSackmann/tennis_slam_pointbypoint>