

An Analysis of Advantage Factors in Men's Tennis – Handedness, Age, Height, and Rank

Shreyes Sridhara

Stonehill International School, Tarahunise Post, Bangalore, Karnataka (KA), 562157, India; shreyesXsridhara@gmail.com

ABSTRACT: This study takes the discussion of the handedness advantage in men's tennis further and analyzes whether it has increased or decreased over the years as playing styles, racket technology, training, and diets of tennis athletes have changed, and as new players have entered the circuit. The results confirm that this advantage favoring left-handed tennis players has indeed diminished over the years to favor right-handers, and this shift was more conspicuous in the highest tier of men's professional tennis than it was in the second and third tiers. I found that the lower the tennis tier, the lower the probability of a shift in the handedness-related advantage over time. In addition to this, other features like age, height, and rank of competing players were investigated to observe whether they too displayed a trend or another advantage. In the higher tennis tiers, younger players had the advantage over their older counterparts while, quite surprisingly, the opposite was true in the lowest tier. Height was only a significant match-determinant in the lower tiers. Furthermore, higher-ranked tennis players were more likely to win against lower-ranked players in descending order of tier.

KEYWORDS: Robotics and Intelligent Machines; Machine Learning; Logistic Regression; Sports Analytics; Left-handedness in ATP.

■ Introduction

In Open Era tennis (began in 1968), it has long been conjectured that there is a handedness-related advantage – that being left-handed is a match determinant in the interactive sport. This claim has been backed by numerous studies. For example, one study by Hagemann divided 54 left-handed and right-handed players from different domains of tennis expertise into three groups and asked them “to predict the subsequent direction of an opponent's temporally occluded tennis strokes on a computer screen”.¹ The results demonstrated that regardless of skill, both right-handed and left-handed tennis players were better at predicting the trajectory of strokes coming from right-handed tennis players. Another study proved the handedness advantage through statistical analysis of a database of tennis matches.² Left-handers comprise approximately 10% to 13% of the world's population; however, they seem to be over-represented in interactive sports, such as tennis and combat sports like boxing and wrestling.^{3,4} Hagemann and Faurie & Raymond suggest two hypotheses to explain the reason behind this overrepresentation seen in tennis: (a) the innate superiority hypothesis and (b) the negative frequency hypothesis.^{1,5}

The innate superiority hypothesis elaborates on the neuropsychological advantages that left-handers possess. For instance, according to Geschwind and Galaburda, regions present in the right hemisphere of the brain are larger for left-handers, serving as an explanation for why left-handers might be better than right-handers at performing tasks that require the right side of the brain (such as attentional and spatial imagery tasks).^{1,6} In addition to this, the results of other studies assert that left-handers possess better anticipation skills and a quicker reaction time with their left hand than right-handers do.⁷⁻¹⁰ Furthermore, as a result of the weaker lateralization

of the brain hemispheres for left-handers, “the stronger non-dominant side or the superior interplay between the two sides could then lead to better overall performance” in motor tasks that demand both hands (such as playing tennis with a double-handed backhand).¹

Referred to by numerous research papers as the negative frequency hypothesis,¹⁻⁵ this hypothesis states that since left-handers are a minority in the world population, left-handed tennis players will play a more significant number of matches against right-handed tennis players, while right-handed tennis players will encounter more tennis players of the same handedness.^{11,12} Therefore, the lack of exposure to the playing style of a left-hander takes the right-handed player by surprise when the latter encounters a left-hander. The ball coming from the opposite side of the court, the topspin on the ball, and the ball's trajectory as it moves through the air all take some time for a right-hander to adjust to. In her autobiography, Monica Seles, a former world number one tennis player on the WTA tour, wrote, “It's strange to play a lefty because everything is opposite and it takes a while to get used to the switch. By the time I feel comfortable, the match is usually over”.² Thus, the motor skills of right-handed tennis players required to address the attack coming from the opposite side of the court are underdeveloped and unpretentious, prompting weaker and less effective responses to these attacks. The notion of negative frequency demonstrates how left-handers can also capitalize on “the movement information” and “anticipate an opponent's intentions” earlier and better than right-handers can, allowing them to distinguish the groundstrokes, the racket speed, and the ball trajectory when it comes off the racket of right-handed tennis players more accurately than right-handed tennis players can distinguish these features from left-handed tennis players.¹

It is evident from the studies conducted (that elaborate on and confirm the existence of a handedness-related advantage for left-handers in an interactive sport like tennis) and the prior research on the topic of handedness and its impact as a match determinant that this is a topic that has been thoroughly addressed. However, I wish to add value to this discussion by shedding new light on and analyzing whether the influence of handedness as a match determinant has changed over the years.

Handedness: I hypothesize that the influence of handedness as a match determinant has waned over the time period I am considering (Open Era tennis) in the highest tier of men's tennis – ATP (Association of Tennis Professionals); on the contrary, I believe that it still remains a prevalent and prominent factor in lower-tier tournaments, such as ATP Challengers (a tier below the ATP) and Futures tournaments on the ITF tour (International Tennis Federation and the third tier of tennis competition). The lower the tier, the lower the probability of a shift in the handedness-related advantage over time.

My theory is predominantly based on the negative frequency hypothesis. In lower-tier tournaments, left-handed tennis players should win more matches against their right-handed counterparts due to right-handers not having much experience playing against lefties to be able to consistently anticipate the left-hander's movements and the ball trajectory when it comes from the opposite side of the court. In addition to this, new tennis players (with limited tennis-related strengths) enter the circuit all the time allowing left-handers to constantly be at an advantage. Hence, in these situations, left-handedness proves to be a prominent match determinant. However, in the ATP level tournaments, the right-handers have played a number of matches against left-handers to reach the level at which they are at and may even train purposefully with opposite-handed players on a regular basis. This means that at the highest level of the sport, tennis becomes more of a game of skills rather than one of advantages; players that enter this circuit must already have a strong and highly-developed game. Also, interestingly, an analysis of the dataset utilized in this study revealed that since the 2000s, the percentage of left-handers on the ATP Tour has been about 12.3%, while this number increases to 13.8% on the Challenger Tour and 12.8% on the Futures Tour. The hypothesis is also in line with the argument that Breznik makes – that the advantage of left-handedness as a match determinant is more noticeable “in the less important types of professional tournament, Challengers and Futures tournaments”.²

I also wanted to investigate whether characteristics like age, height, and rank showed a trend over the years that I could interpret as an advantage.

Age: I expect juvenile tennis players to perform better than older players due to their better fitness and energy levels.

Height: I believe that tall tennis players have an advantage over their shorter peers since their height should help them serve at faster speeds (as they have more body mass), and at sharper angles (giving them a bigger margin for error).

Rank: Finally, higher-ranked players are believed to be more successful against their lower-ranked rivals because they must

be better tennis players (to have reached the rank at which they are at). It is to be noted that all of my claims on age, height, and rank are based on priori reasoning.

The goal of the study is to validate my main hypothesis based on handedness and my other claims on age, height, and rank statistically by constructing a machine learning model over a dataset of tennis matches and then interpreting the results.

■ Methods

Database:

The data utilized for this study was obtained from Jeff Sackmann's “tennis_atp” GitHub repository.¹³ The repository contains datasets of tennis stats, ranking, and results of men's singles tennis matches from different tiers – ATP Tour, ATP Challenger Tour, ITF Futures Tour – from 1968 to 2020. Each dataset comprises all the matches that occurred for each year in the respective tier along with 49 features, including the handedness, the age, the height, and the ranking of both winner and loser, the surface played on, the draw size of tournaments, and the match date.

For this analysis, I considered the handedness (principal), the age, the height, and the rank of the competing tennis players and investigated how the coefficients of these variables changed with time. The other features in the datasets, like the draw size and the match date, were rejected since they wouldn't have been helpful for modeling the winner. The features I chose are also displayed alongside tennis players on the ATP official website and those that have been speculated to affect the result of a tennis match.

Materials and Tools:

I performed the modeling and the analysis using Python. Python is a general-purpose, high-level object-oriented, programming language and one of the benefits of this programming language is the variety of libraries that can be accessed through the language, such as pandas, matplotlib, and sklearn. I used pandas for data manipulation, sklearn to train and test the logistic regression model and make further predictions using the model, and matplotlib to depict the results obtained through the model in a more perceptible manner.

Modelling Handedness:

I used machine learning to train models from the vast dataset and analyze the results. Machine learning algorithms (systems of instructions or procedures to solve problems) build models from data. In today's modern world, machine learning is used in built speech recognition software on smartphones, product recommendations, social media, and cybersecurity to prevent online fraud. For this study, I used supervised learning so that errors with the output of the model can be compared to the expected output data, thereby adjusting and improving the model. In supervised learning, an algorithm is provided with labeled examples or training data. Subsequently, the program learns how to map the input data supplied to the desired output to then be able to make predictions of similar future events.¹⁴ This study is an example of a binary classification problem as there are only two possible outcomes – loss or win; the intention is to use handedness, age, height, and rank to predict which player will either win or lose.

Using Logistic Regression :

I used logistic regression since this was a binary classification problem. In logistic regression one or a set of predictor variables ('x') are used to estimate a binary output variable ('y'); simply put, it will help us estimate the probability of label 'y' when a feature 'x' is inputted. In my analysis, I computed the differences in handedness, age, height, and rank between the two competing players of each match. Table 1 below summarizes this process.

Table 1: Table demonstrating the difference method used to obtain the 4 predictor variables. The significance of a positive coefficient for each variable is also shown.

| Determinant | Difference Between | Significance of a positive coefficient |
|-------------|---|---|
| Handedness | right (which is taken as 1) & left (which is taken as 0) | Right-handed tennis players have an advantage |
| Age | older player age & younger player age | Older tennis players have an advantage |
| Height | taller player height & shorter player height | Taller tennis players have an advantage |
| Rank | lower ranked player rank (larger number) & higher ranked player rank (smaller number) | Lower-ranked tennis players have an advantage |

I then used these differences as the predictor variables to estimate the probability of a player with a particular feature winning.

Before building the model, I had to make a few modifications to the data frames. I replaced the missing values with the mean value of the respective columns and normalized the difference columns. These were done to ensure I didn't have NaN values in the dataset before I computed the differences in the features and implemented my logistic regression model.

In logistic regression, the beta coefficients (represented using the symbol 'β') of the equation inform us by how much the outcome variable changes with every unit change in the predictor variable, provided that the other variables in the equation are held constant.¹⁵ Positive coefficients imply a higher probability of the event occurring, negative coefficients imply the opposite, and coefficients near zero signify that the feature is of less significance.

Logistic regression uses the logit function, which can be represented as

$$\text{logit}(p) \text{ or } \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \dots + \beta_kx_k$$

Since I only consider four features in this study, my logit function will be

where p is the probability of the winner label occurring

$$\ln\left(\frac{p}{1-p}\right) = \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4$$

(win or loss), x_1 represents the difference in handedness, x_2 represents the difference in age, x_3 represents the difference in height, and x_4 represents the difference in rank; $\beta_1, \beta_2, \beta_3, \beta_4$ are their respective coefficients. On a further note, it is important to understand that when focusing on a certain predictor variable – for example, the difference in age of the two competing players – I am keeping every other feature, including random effects like luck in a tennis match, constant.

To name a few such random effects, sometimes net cord balls drop in and bad calls made by umpires require the point to be replayed when clearly one player was in control of the point. This means that (using the same example) I am only trying to see the effect of being a year older or younger (the two players are hypothetically of the same handedness, height and rank) on the result of the match. Therefore, the differences in handedness, height and rank are 0, which also means that I consider the two of them to have equal amounts of luck by taking the constant term as 0. This paper also does not capture the impacts of such random effects on a match; hence, β_0 is always taken as 0 and can therefore be omitted. Furthermore, the logistic regression model was run with every match between opponents of opposite handedness, and adding a constant term would be similar to a priori reasoning that one of the players in the match-up has an advantage over the other from the get-go in a case where every other variable is controlled. I also randomized the side of the court and so, by symmetry, β_0 will be 0 once again.

Procedure:

Since I only considered matches with opposite-handed tennis players (a right-hander and a left-hander), the other handedness combinations, including the 'U' values for handedness (the undefined or unknown values for handedness in the datasets) were dropped.

I discovered that on the ATP Tour (Tier 1), on average, left-handed tennis players won 52.4% of the time when playing against right-handers; this percentage drops to 51.4% for Challengers (Tier 2) and Futures (Tier 3).

The breakdown of the matches above shows that left-handed tennis players won a greater percentage of matches against the more prevalent right-handed players on average. These results confirm the work of other prior studies that showed a handedness-related advantage exists in tennis, even at the highest level of the sport.¹⁻⁵ However, I was curious if the influence of this handedness-related advantage on the different tennis circuits changed over time.

Statistical Analysis using Scikit-learn (sklearn):

I ran models for every year in each tennis tier and plotted the output variable coefficients of the respective features over the time period considered. Additionally, I performed loess smoothing on all curves to help see the trends and relationships better.

Results and Discussion

Handedness:

In Figure 1, the general trend of the ATP shows that left-handers had an advantage in the 1970s until the 1990s; now right-handed tennis players do. On the Challenger Tour, the handedness advantage prevailed during the 1980s but became less and less important afterward. A quick analysis of the dataset explains a possible reason behind this. The percentage of left-handed tennis players on the Challenger Tour went from 14.6% at the start of the 80s to an astounding 21.6% at the end of the decade. Therefore, according to the negative frequency hypothesis, right-handed tennis players were challenged more often by left-handed tennis players as the decade progressed, and this constant exposure would have helped

them get used to the different trajectories and the top spin that left-handers induce on the ball. Thus, the left-handedness advantage diminished during this decade.

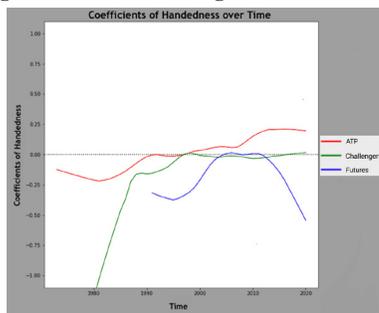


Figure 1: Graph of the coefficients of right-handedness on the ATP Tour, Challenger Tour, and Futures Tour over time.

Handedness seems to have had the least importance on the Challenger Tour over the last two decades, suggesting that it might be an even fight between players of different handedness. This once again can be explained by the fact that the percentage of left-handed tennis players on the Challenger Tour since the 1990s has stayed between 20% and 21%; hence the exposure to left-handers has more or less stayed the same on the tour.

In the Futures, handedness was important in the 1990s and in the 2010s (and continues to follow an increasing trend) but hardly showed any relevance during the middle years. Over the course of the last decade, the percentage of left handers on the Futures Tour fell from 13.6% to 11.9%. Therefore, in line with the negative frequency hypothesis, right-handers on the Futures Tour would have had limited exposure to playing tier level matches against their left-handed counterparts, creating a left-handed advantage.

I can deduce that currently, handedness is a salient factor in the Futures and in the ATP: the left-handedness advantage continues to increase on the Futures, while right-handedness prevails in the ATP. The handedness-related advantage did not shift over the period considered on the Challenger Tour and Futures Tour (lower-tier tournaments); however, this observation was more significant on the Futures graph. This aligns with my theory: the lower the tier, the lower the probability of a shift in the handedness-related advantage over time.

Figure 1 is in support of the hypothesis that the influence of handedness as a match determinant has waned over the time period considered on the ATP Tour – the highest tier of men’s professional tennis. The coefficient did not go below zero at all during the last two decades, demonstrating that the leftie advantage was no longer as influential or prevalent in the top tier of tennis as it was during the 90s, the 80s, or the 70s. I can confirm that right-handed tennis players had a greater probability of winning matches after the 21st century compared to the years before using the logit function.

In order to compare two players of different handedness, I controlled the age, the height, and the rank by keeping them constant. As a result, the input variables other than the difference in handedness (difference in age, difference in height, difference in rank of the two players) were zero.

$$\ln\left(\frac{p}{1-p}\right) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$$

$$\begin{aligned} \ln\left(\frac{p}{1-p}\right) &= \beta_1 * (\text{difference in handedness}) + \beta_2 * (\text{difference in age}) + \beta_3 \\ &\quad * (\text{difference in height}) + \beta_4 * (\text{difference in rank}) \\ &= \beta_1 * (1) + \beta_2 * (0) + \beta_3 * (0) + \beta_4 * (0) \end{aligned}$$

$$\ln\left(\frac{p}{1-p}\right) = \beta_1$$

Hence,

For handedness, the difference was always one.

When one of the beta coefficients of the difference in handedness from the results is entered, it gives us the probability of a certain handedness (either right or left depending on whether the coefficient is positive or negative) winning.

For instance, when I consider two players with the same age, height, and rank, and take the beta coefficient of 0.2 (which

$$\ln\left(\frac{p}{1-p}\right) = \beta_1$$

$$\ln\left(\frac{p}{1-p}\right) = 0.2$$

$$\therefore p = 0.550$$

lies on the ATP curve after the 2000s),

According to the model, right-handers had a 55% chance of winning against lefties.

On the other hand, when I take the coefficient -0.2 (which lies on the ATP curve between 1970 and 1990)

$$p=0.450$$

We see that right-handed tennis players only had a 45% chance of winning.

At the start of the 1980s, on the Challenger Tour, left-handed tennis players had a massive advantage over right-handers.

$$p=0.269$$

Left-handed tennis players had a 73.1% chance of winning.

The curve representing the Futures Tour was never over the x-axis for a significant amount of time, therefore, suggesting that the probability of winning would be in favor of the left-handers. The coefficients between (approximately) 2003 and 2015 illustrate that the advantage had lessened during this time period compared to previous and future years; however, the coefficients post 2015 and prior to 2000 were large negative values and demonstrate a convincing left-handed advantage.

An intriguing matter to investigate is how prevalent the left-handed advantage would be if Nadal’s match data is removed from the data set. Nadal has spent more than 800 weeks in the top 10, and continues to beat top right-handed players. For instance, Roger Federer is considered as one of the best players against left-handers, and poses a 125-36 (win-loss) record against lefties on the ATP Tour. However, 23 of those 36 losses have come just against Nadal.

It’s possible that Nadal’s dominance on the ATP Tour skews the left-handed advantage in its favor. But by exactly how much?

Upon removing Nadal's record and considering the time frame in which he has played,

$$\ln\left(\frac{p}{1-p}\right) = 0.4$$

$$p=0.60$$

According to the model, right-handers would have had a 60% chance of winning against lefties, if Nadal never played on ATP tour. This demonstrates that without Nadal, the right-handed advantage on the ATP Tour increases by 5%: being left-handed, is a clear disadvantage in this case.

Similarly, to find the chance of winning with the other coefficients, I kept the remaining features of the two players constant and followed the same process.

Age:

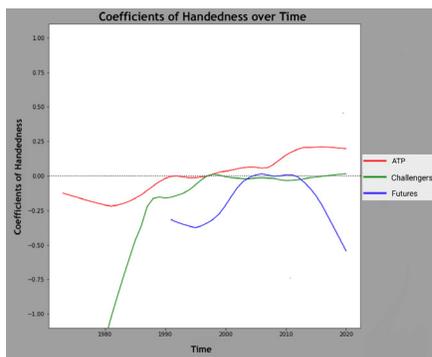


Figure 2: Graph of the coefficients of age on the ATP Tour, Challenger Tour, and Futures Tour over time.

Figure 2 shows that age has different impacts on the tiers. The ATP and Challengers graphs align with my claim that younger tennis players tend to perform better and defeat the older players. This juvenile advantage was more pronounced on the Challenger Tour than on the ATP Tour only prior to 1990. Post 1990, younger tennis players had a higher probability of winning on the ATP Tour than on the Challenger Tour. By contrast, older tennis players were better on the Futures Tour circuit, but this advantage favoring older players has decreased over the last couple of years. Older tennis players are better on the Futures Tour since they are likely to have more match experience. The level of tennis on the Futures Tour is not as great as that on the ATP Tour since it is the lowest rung of men's professional tennis. Often, matches involve taking advantage of the opponent's weakness. For instance, some players might have a weaker backhand compared to their forehand and their opponents will simply target their backhand side throughout the match. Older players are more likely to have had more time to work on bettering their overall game, giving these men an edge over the young teenagers who usually enter Futures tournaments. However, the last few years have seen the likes of Felix Auger Aliassime and Jannik Sinner; prodigies like these two have beat numerous older players in the lower tier tournaments like the Futures, before they advanced to the ATP Tour. On the contrary, to even be on the ATP Tour, players must have almost no weaknesses. At that level, the game becomes more of a mental, physical, and tactical sport rather than one that involves targeting weaknesses.

When I consider two players who are one year apart in age, I get the normalized difference in age from my training set to be 0.0272.

On the Challenger Tour during the 1980s,
 $p=0.483$

The chance of winning when being one-year younger was 51.7%.

However, on the Challenger Tour during the 2010s,
 $p=0.501$

It shows that the chance of winning for a younger player was 49.9%. The numbers show how the chance of winning for younger players and therefore their advantage diminished over time on the Challenger Tour.

On the ATP tour, 55 of the last 64 Grand Slams have been won either by Roger Federer, Novak Djokovic, or Rafael Nadal, who are collectively known as "The Big Three". These three tennis players have continued to dominate the top tournaments despite the numerous young talents that have emerged over the last few years and despite the fact that they are all in their 30s (Federer will turn 40 in August 2021). Over the last couple of years, the age-related advantage has not only reduced on this tour, but has also become positive – in support of older players.

Building the same logistic regression model without "The Big Three" in the data set would reveal the extent to which the previous model had been biased towards older tennis players in recent years on the ATP Tour.

$$p=0.489$$

Interestingly, the chance of winning for a younger player would have been 51.1% if "The Big Three" didn't compete on the ATP Tour. Although not a salient advantage, the age-related advantage would have been in favor of younger players over older players.

Height:

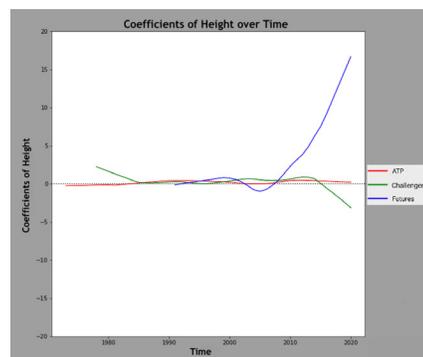


Figure 3: Graph of the coefficients of height on the ATP Tour, Challenger Tour, and Futures Tour over time.

Figure 3 demonstrates that taller tennis players on the Challenger circuit had a slightly higher probability of winning against shorter players in the late 1970s and early 1980s. Now, shorter players have the advantage. On the ATP tour, no certain height-related advantage could be deciphered as the curve constantly hugged the x-axis, suggesting that height hasn't had much importance on this tour. On the other hand, being taller than the opponent on the Futures Tour was a significant advantage. The dataset reveals that on the Futures Tour, the mean winning player height was 182.6 cm during

the 90s as compared to 188.1 cm during the 2010s. Taller tennis players on this tour serve better and faster for the following reasons: their heights provide a larger margin for error and a greater angle to serve out wide. In addition to this, they are more likely to exert a larger force on the ball (more body mass, which correlates to a greater change in momentum). The lack of experience in playing against taller players means that height is a significant match determinant on the Futures Tour. This puts shorter tennis players, who have a smaller wingspan and racket-arm extension, on the Futures Tour at a huge disadvantage. And since the mean height of tennis players has increased on this tour, this advantage has surged.

When I take two players who only differ by 1 cm in height, I get the normalized difference in height from my training set to be 0.0349. This predictor variable was plugged into the logit function.

During the 1980s on the Challenger Tour,
 $p=0.517$

This implies that the chance of winning for a player who was a centimeter taller than the opponent was 51.7% during the 1980s.

And as stated above, this chance of winning shifted to favor the shorter players during the last couple of years.

$p=0.474$

A shorter player on the Challengers had a 52.6% chance of winning.

On the Futures tour, taller players had a significantly higher chance of winning than on the other two tiers during the 2010s.

$p=0.612$

A 61.2% chance of winning for a taller player.

Rank:

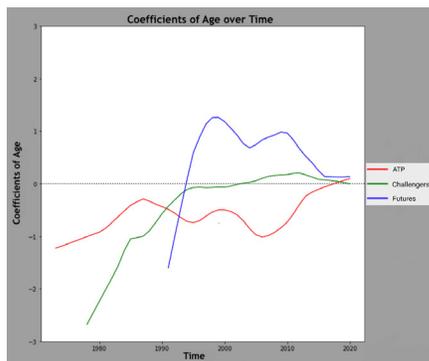


Figure 4: Graph of the coefficients of rank on the ATP Tour, Challenger Tour, and Futures Tour over time.

In Figure 4, all three curves on the graph show that higher-ranked tennis players have always had the advantage in a tennis match when playing against a lower-ranked opponent, supporting my claim that higher-ranked players had the upper hand between the two. Between the 1970s and the early 1980s, the probability of higher-ranked tennis players being more successful against their lower-ranked counterparts decreased on the ATP tour. Around the mid-1980s, the curves for both the ATP and Challengers started to plateau, suggesting that the advantage has had the same importance on the respective tours for a while. However, the ATP curve plateaued

lower than the Challenger curve implying that the advantage is more prominent on the ATP Tour and that the probability of a higher-ranked player defeating a lower-ranked player is higher on the ATP Tour than on the Challenger Tour. Since the Futures curve is above the other two and has a general negative trend, it implies that upsets are more likely to happen on the ITF Futures Tour and that this advantage is less salient on this tour compared to the other two circuits; however, the negative trend suggests that it is slowly becoming more relevant and important.

To quantify the advantage seen above, I chose two players who only differ by 1 ranking position, and got the normalized difference in height from my training set to be 0.111.

On the Futures Tour during the 1990s,
 $p=0.418$

When a player faced an opponent who was 1 rank below him during the 1990s, he had a 58.2% chance of winning against his opponent. I also know from the graph that the advantage of being higher-ranked in the Futures slowly increased with time; this can be confirmed by taking a coefficient from the 2010s.

$p=0.315$

When a player faced an opponent who was 1 rank below him during the 2010s, he had a 68.5% chance of beating his opponent.

■ Conclusion

I study how the handedness-related advantage develops over time on three consecutive tiers of men's tennis. The results agree that the influence of handedness as a match determinant has waned over time on the ATP Tour, but is still a prevalent factor on the ITF Futures Tour, while the ATP Challenger Tour seems to be an unbiased ground now.

This study substantiates my belief that younger players have a higher probability of beating older players. This claim holds true on the ATP Tour and Challenger Tour, but not on the Futures Tour. With rank, I confirm that higher-ranked players perform better, and that lower-ranked players have a better chance at winning against their counterparts on the Futures Tour than on the other two tiers. With height, the Futures Tour shows that this feature was important; on the contrary, shorter players have the advantage in Challenger matches.

A limitation of the study is the incomplete datasets on the GitHub repository during the first couple of years of Open Era tennis. The unavailability of data meant I couldn't get to use the exact values for those four features for my logistic regression model. I had to settle with the mean values of the columns they belonged to instead. Future work could undertake the same approach in women's, doubles, and mixed doubles tennis and their respective tiers; the findings should then be compared to the ones in this study.

Implications: From a tennis player's perspective:

As a tennis player and a right-hander who has played the sport for the last 8 years, I have first-hand experience of playing against left-handers and also evidence to support my main hypothesis. Having played low-tier tennis tournaments, such as AITA (All India Tennis Association), I can say that lefties have an advantage in this circuit, once again due to the negative frequency hypothesis. Righties like myself (who don't get

a chance to practice or play with left-handers often) struggle against left-handed tennis players during these tournaments due to lack of experience of playing against lefties and due to psychological influences (it's common for tennis coaches to constantly label left-handers as difficult opponents). More often than not, the left-handed tennis player will control and dictate the majority of points in such a matchup, and will end up winning the match.

■ Acknowledgements

I would like to thank my mentor Patrick Chao, who is a statistics Ph.D. student at the University of Pennsylvania, for his guidance during the research and writing processes. Furthermore, I would like to thank the faculty at Lumiere Education for providing me with the resources to make this project successful.

■ References

1. Hagemann, N. The advantage of being left-handed in interactive sports. <https://link.springer.com/article/10.3758/APP.71.7.1641> (accessed Apr 21, 2021).
2. Breznik, K. On the gender effects of handedness in professional tennis. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3761843/> (p.72) (accessed Apr 21, 2021).
3. AP, R. M. P. D. D. A. B. M. Frequency-dependent maintenance of left handedness in humans. <https://pubmed.ncbi.nlm.nih.gov/9025310/> (accessed Apr 21, 2021).
4. Richardson, T.; Gilman, R. T. Left-handedness is associated with greater fighting success in humans. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6817864/> (accessed Apr 21, 2021).
5. Faurie, C.; Raymond, M. Handedness, homicide and negative frequency-dependent selection. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1634940/> (accessed Apr 21, 2021).
6. Norman Geschwind, M. D. Cerebral Lateralization. <https://jamanetwork.com/journals/jamaneurology/article-abstract/584161> (accessed Apr 21, 2021).
7. PMC, E. <http://europepmc.org/article/med/11104547> (accessed Apr 21, 2021).
8. P., B. S. B. Manual reaction time asymmetries in human subjects: the role of movement planning and attention. <https://pubmed.ncbi.nlm.nih.gov/11711210/> (accessed Apr 21, 2021).
9. APA PsycNet. <https://psycnet.apa.org/record/2006-03817-005> (accessed Apr 21, 2021).
10. Vasconcelos, O.; Barreiros, J.; Barbosa, R.; Trifilio, F. Functional asymmetry in a simple coincidence-anticipation task: Effects of handedness. <https://www.tandfonline.com/doi/abs/10.1080/17461390802603903> (accessed Apr 22, 2021).
11. Akpınar, S.; Bicer, B. Why Left-handers/footers are overrepresented in some sports? . http://mjssm.me/clanci/MJSSM_Sept_2014_Akpınar_33-38.pdf (accessed Oct 2020).
12. Loffing, F.; Hagemann, N.; Strauss, B. Left-Handedness in Professional and Amateur Tennis. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0049325> (accessed Apr 21, 2021).
13. JeffSackmann. JeffSackmann/tennis_atp. https://github.com/JeffSackmann/tennis_atp (accessed Apr 22, 2021).
14. Joseph A. Cruz, D. S. W. Applications of Machine Learning in Cancer Prediction and Prognosis - Joseph A. Cruz, David S. Wishart, 2006. <https://journals.sagepub.com/doi/full/10.1177/117693510600200030> (accessed Apr 21, 2021).
15. Shewchuk, J. R. Concise Machine Learning. <http://people.eecs.berkeley.edu/~jrs/papers/machlearn.pdf> (accessed Oct 2020).

■ Author

Shreyes Sridhara is an IB student in his junior year at the Stonehill International School in Bangalore, India. He enjoys using statistical analysis to analyze and extract practical information from large-scale data, and hopes to major in Computer Science in college.