

MCM/C

February 3, 2024

Contents

1	Background Information	1
1.1	Some Info on Tennis Scoring	2
2	Who Has A Bigger Momentum? A Visualisation Approach	3
3	Does Swings Happen Randomly?	5

1 Background Information

The 2023 Wimbledon Championships marked a significant event in the world of tennis, particularly in the Gentlemen's Singles category. One of the most notable matches was the final between the young Spanish sensation, 20-year-old Carlos Alcaraz, and the seasoned champion, 36-year-old Novak Djokovic. In a surprising turn of events, Alcaraz defeated Djokovic, ending the latter's impressive winning streak at Wimbledon since 2013. This match stood out not just for its outcome but also for its dramatic shifts in momentum, a concept often discussed in sports but hard to quantify. Throughout the match, both players exhibited periods of dominance, with the advantage oscillating between them in an unpredictable manner. This intriguing match, along with other games from the tournament, provides a rich dataset to explore the elusive concept of momentum in tennis and its impact on match outcomes.

Momentum in sports has been studied by many scholars in various fields. Some aim to quantify momentum from a statistical point of view, while others try to explain it from a sports psychological point of view. All fields are equally valid. However, our team will build a mathematical model to explain and answer the following questions:

- Create a model to capture the flow of play in tennis matches, identifying the player performing better at any given time and the extent of their advantage.
- Provide a visualization based on your model to depict the match flow.

- Use our model to evaluate the claim: "swings in a tennis match are random, as opposed to being influenced by momentum."
- Develop a model that predicts when the flow of play is about to change. Identify factors that might be related to these shifts.
- Test your model on other matches to evaluate its effectiveness and generalizability.
- Produce a report (max 25 pages) with your findings. Include a memo summarizing results and advice for coaches regarding momentum and preparation for matches.

1.1 Some Info on Tennis Scoring

Game

- Each point won counts as one.
- The first to score 4 points wins the game.
- At 3 points each, the score is 'Deuce.' From Deuce, a player must win by two points to win the game.
- In international tennis, the scores of 0, 1, 2, and 3 points are represented by the English words Love, 15, 30, and 40, respectively.

Set

- The first player to win 6 games wins the set.
- If both players win 5 games each, the set is won by the first player to lead by two games.

Tie-break Scoring When the game score in a set reaches 6-all, one of the following tie-break methods is used:

- Long set: One player must lead by two games to win the set.
- Tie-break (or 'short set'): Except in the final set (unless otherwise specified), the following rules apply:
 - The first player to score 7 points wins the game and the set (at 6-all, a player must win by two points).
 - The first server serves the first point; thereafter, players alternate serving two consecutive points.
 - The first point is served from the right court, the second from the left, and the third from the right.
 - Players change ends after every six points and at the end of the tie-break.

Best of Five Sets Format The match is won by the player who first wins three out of five sets.

2 Who Has A Bigger Momentum? A Visualisation Approach

Before we begins any real development of our model, we need to visualise. Visualisation provides a clear and graphical approach. But, we need to state our definition for "momentum". In one of the academic papers we found, the authors [1] identified momentum as:

Definition 1 (Momentum by Chen et al.). *Score difference in a given period of time.*

In the paper. Momentum is defined mathematically as:

$$M(t, t_0, \gamma) = \begin{cases} y(t_0 + t) - y(t) & \text{if } y(t_0 + t) - y(t) > \gamma \\ 0 & \text{otherwise} \end{cases},$$

where $y(t)$ represents the score difference between the home and visiting team at time t_0, t denotes an increment of game time and γ represents the threshold value of momentum. Similarly, the momentum of the other team is

$$M(t, t_0, \gamma) = \begin{cases} y(t_0 + t) - y(t) & \text{if } y(t_0 + t) - y(t) < -\gamma \\ 0 & \text{otherwise} \end{cases}.$$

Equipped with this definition, we now proceed to visualisation.

Data Preprocessing First we need to process some data in the "Wimbledon_featured_matches.csv".

- Convert "match_id" into unique sequential numbering
- Convert all time to seconds
- Standarise all scoring into points(e.g. 1,2,3). Reason being nonuniform scoring could affect calculation

Now, based on your definition of momentum, we have generated the following graphs:



Figure 1: placeholder

We want a model that accounts for more than just score difference. We made 2 considerable additions to the definition of "momentum".

- Added in "server advantage". Numerous papers have stated that server advantage is an important psychological effect to increase the probability of scoring [3] [4], though it alone does not determine the competition outcome. **We model the effect by adding a "server scoring probability" to whomever is serving.**
- Added in "ace". Ace means a server wins a point by serving. This significantly boosts the server's morale.

A modified visualisation shows:



Figure 2: placeholder

3 Does Swings Happen Randomly?

The study of momentum in sports generated great interests among scholars, sports coaches and athletes. The debate is still ongoing—some people do not believe in momentum [2], just like our dear tennis coach.

It was mentioned earlier that the trends of momentum rising and falling reflect performance. We can define the turning point of the rising and falling trends as the inflection point. In each game, by linearly fitting the overall performance of player 1 in the game with the number of balls scored, we can determine the slope that reflects the rising and falling trends. When the performance changes (i.e., when the sign of the slope changes), an inflection point occurs. To filter out the inflection points where the trend changes slowly, only $|\text{slope}|$ above a certain threshold are considered as inflection points. These points are marked with red vertical lines on the graph.

In this scenario, we aim to develop a mathematical model to identify turning points in a player's performance trend during a game. The turning points are defined as moments when the trend of performance changes from increasing to decreasing or vice versa, determined by the slope of a fitted line.

- Data Representation
 - Let t_i represent time in the game.
 - Let y_i represent the player's score at each t_i .
- Linear Regression Model

- We use a simple linear regression model to fit the data points (t_i, y_i) .
- The linear model is given by $y = mt + b$, where m is the slope and b is the y-intercept. The slope m indicates the trend of the player's performance. A positive slope implies an increasing trend, while a negative slope indicates a decreasing trend.
- To find the best-fit line, we minimize the sum of squared differences between the actual performance scores and the scores predicted by the linear model. The objective is to minimize the Mean Squared Error (MSE), given by:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - (mt_i + b))^2$$

- Identifying Turning Points

- A turning point is identified when there is a change in the sign of the slope (from positive to negative or vice versa).
- Moreover, to filter out minor fluctuations, a threshold is set for the slope. Only when the absolute value of the slope changes and is greater than a specified threshold (in this case, 0.15) is a turning point acknowledged.
- Mathematically, if m_{prev} and $m_{current}$ are the slopes of two consecutive segments and $|m_{current}| > 0.15$, a turning point is identified when $\text{sign}(m_{prev}) \neq \text{sign}(m_{current})$.

Now we have identified all turning points that are actually swings, we need to test for their randomness. We have imposed certain conditions on turning points and the randomness test is as follows:

Let's denote:

1. N as the total number of games.
2. n_1 as the number of games with turning points.
3. n_2 as the number of games without turning points.
4. $N = n_1 + n_2$.
5. R as the total number of runs (a run is a sequence of consecutive games either all with turning points or all without turning points).

The expected number of runs ($E(R)$) and the variance of the number of runs ($Var(R)$) are given by:

1.

$$E(R) = \frac{2n_1n_2}{N} + 1$$

2.

$$Var(R) = \frac{2n_1n_2(2n_1n_2 - N)}{N^2(N - 1)}$$

The test statistic (Z) is then calculated as:

$$Z = \frac{R - E(R)}{\sqrt{Var(R)}}$$

While P-value is given by:

$$\text{p-value} = 2P(Z > |z|)$$

Interpretation

- If the value of Z is significantly high or low, it indicates that the turning points are not randomly distributed, suggesting a pattern or trend in the player's performance changes.
- A high p-value (usually > 0.05) implies that the turning points are randomly distributed, indicating no specific pattern in performance changes.

Results We have obtained very low p-values, which suggests that swings are not random.

References

- [1] Tao Chen, Qingliang Fan, Kai Liu, and Lingshan Le. Identifying Key Factors in Momentum in Basketball Games. *Journal of Applied Statistics*, 48(16):3116–3129, 7 2020.
- [2] David Hale. Is momentum real? an in-depth investigation of sports' most overused term - espn. *ESPN.com*, December 2021.
- [3] Jan R. Magnus Klaassen and Franc J. G. M. On the advantage of serving first in a tennis set: Four years at wimbledon. *Journal of the Royal Statistical Society. Series D*, 48(2):247–256, 1999.
- [4] I. M. MacPhee, Jonathan Rougier Pollard, and G. H. Server advantage in tennis matches. *Journal of Applied Probability*, 41(4):1182–1186, 2004.