

## Swings in Tennis Match: The Impact of Momentum on Flow of Play

### Summary

The striking swings observed in tennis matches are often ascribed to "momentum", which can significantly boost a player's confidence. To delve into momentum's influence on match dynamics and its effect on the flow of play, we utilize data from the 2023 Wimbledon Gentlemen's tournament for modeling. Prior to model development, we execute data cleaning, correct format errors, address missing and outlier values.

Initially, we develop Model 1 to capture the flow of play, defining "flow" in both a general sense and at specific moments. The flow at a specific point ( $f$ ) is identified by the performance differential between the two players at that moment. The overall flow ( $F$ ) represents the cumulative value of  $f$ . We select qualitative and quantitative indicators from serving, match conditions, scoring, and points lost to construct our functional model. The **Analytic Hierarchy Process (AHP)** is employed to ascertain the weights for each indicator.

Next, we quantify the match momentum as the momentum difference between the two players by **Principal Component Analysis (PCA)**. We calculate **point-biserial correlation coefficient** and **Spearman's rank correlation** for analyze the relationship between swings in play( $SW$ )/runs of success( $Q$ ) and momentum and conduct **run tests** determine if  $SW$  and  $Q$  are random. Acknowledging momentum's lag effect, we theorize that momentum correlates with swings in play and runs of success, with the absence of significant correlations likely due to the cumulative effect not yet reaching a discernible level.

Furthermore, we employ features like cumulative score difference, cumulative movement distance difference, serve success rate difference, and match momentum to develop a **logistic regression model** through time series data windowing and parameter estimation by the **quasi-Newton method**. The model highlights momentum as the foremost predictor of swings (0.446), succeeded by score difference (0.169), achieving an average accuracy of 97.2% across 30 men's tennis matches, with an average absolute error of 0.0234. **Sensitivity analysis**, through adjustment of the test set proportion, demonstrates model stability within half of its parameter range, indicating robustness and reliability.

Moreover, we apply our model to 30 Wimbledon women's singles matches, securing an average prediction accuracy of 95.33%. We also assess the model's efficacy in table tennis matches, where only scoring information is accessible. By revising the model's definition expression to omit error metrics, we achieve an average prediction accuracy of 94.73%.

In the memo to coaches, we explain momentum's cumulative impact, providing advice at five different stages of momentum shift from disadvantageous to advantageous, and suggest psychological and tactical preparations to tackle events that influence the match flow.

**Keywords: Momentum; Tennis Match; AHP; PCA; Correlation Analysis; Hypothesis Test; Logistic Regression**

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
1.1	Problem Background . . . . .	4
1.2	Restatement of the Problem . . . . .	4
1.3	Our Work . . . . .	5
<b>2</b>	<b>Assumptions and Justifications</b>	<b>5</b>
<b>3</b>	<b>Notations and Glossary</b>	<b>6</b>
3.1	Notations . . . . .	6
3.2	Glossary . . . . .	6
<b>4</b>	<b>Data Preprocessing</b>	<b>7</b>
<b>5</b>	<b>Model I: Quantitative Model of Flow and Player Performance Based on AHP to Determine the Indicator Weights</b>	<b>7</b>
5.1	Definition of Flow . . . . .	7
5.2	Definition of Performance . . . . .	8
5.2.1	Indicators Selection . . . . .	8
5.2.2	Comprehensive Evaluation Data Processing . . . . .	9
5.2.3	Mathematical Model Establishment . . . . .	10
5.2.4	Determination of Indicator Weights . . . . .	10
5.3	Analysis and Visualization of Results . . . . .	12
<b>6</b>	<b>Model II: Momentum Evaluation Model Based on PAC and Statistical Analysis</b>	<b>13</b>
6.1	Definition of Momentum in Tennis Matches . . . . .	13
6.1.1	Indicators Selection . . . . .	14
6.1.2	Comprehensive Evaluation Data Processing . . . . .	14
6.1.3	Function Model Establishment . . . . .	14
6.1.4	Determination of Indicator Weights . . . . .	15
6.1.5	Analysis and Visualization of Results . . . . .	16
6.2	Is there a correlation between swings in play and runs of success with momentum?	16

6.2.1	Swings in Play . . . . .	16
6.2.2	Runs of Success . . . . .	17
6.2.3	Correlation Results Analysis . . . . .	17
6.3	Whether "Swings in Play" and "Runs of Success" are Random . . . . .	18
6.4	Assess the Claim of the Coach . . . . .	19
<b>7</b>	<b>Model III: The Swings Prediction Based on Logistic Regression Model Combined with Time Series Analysis</b>	<b>19</b>
7.1	Model Establishment . . . . .	19
7.2	Basic Principles . . . . .	20
7.3	Model Estimation . . . . .	20
7.4	Results Analysis . . . . .	21
7.5	Advise for a Player Going into a New Match against a Different Player . . . . .	21
<b>8</b>	<b>Model Validation and Generalization</b>	<b>21</b>
8.1	Model Validation . . . . .	21
8.1.1	Accuracy Analysis . . . . .	22
8.1.2	Sensitivity Analysis . . . . .	22
8.2	Model Generalization . . . . .	23
<b>9</b>	<b>Model Assessment</b>	<b>23</b>
9.1	Strengths . . . . .	23
9.2	Weaknesses . . . . .	24
<b>10</b>	<b>Memo</b>	<b>24</b>

# 1 Introduction

## 1.1 Problem Background

In the 2023 Wimbledon Gentlemen's final, 20-year-old Spanish rising star Carlos Alcaraz defeated 36-year-old Novak Djokovic. In this particular match, the swings in play were apparent as the advantages shift frequently between the two players. The incredible swings, contributing to dramatic shifts in the player who appear to hold the advantage, which encompass multiple points and even entire games, are commonly associated with the concept “**momentum**”.

**Momentum in sports refers to “strength or force gained by motion or by a series of events” that can influence the outcome of a game or even a match.** During a match or game, it is possible that teams or players sense a surge of the momentum, or “strength/force”, which is challenging to measure. By tracking and analyzing the data of Wimbledon 2023 Gentlemen's singles matches, we may be able to measure the momentum and gain a better understanding of the factors that create or affect momentum during a match.

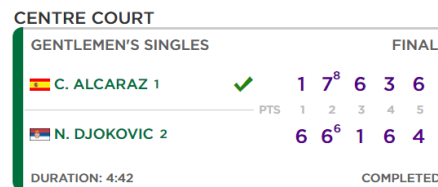


Figure 1: The 2023 Wimbledon Gentlemen's final

## 1.2 Restatement of the Problem

Considering the background information and restricted conditions identified in the problem statement, we need to solve the following problems:

- **Problem 1:** Develop a model that effectively captures the flow of play by monitoring the sequence of point scoring. Utilize this model in one or more matches to discern the player who perform better, providing a quantitative measure of their superior performance at specific points.
- **Problem 2:** Examine the assertion that whether the swings in play and runs of success by one player are random using the above model.
- **Problem 3:** Determine the indicators that unveil an imminent shift of the flow, marking a transition of advantage from one player to another.
  - Develop a model that predicts the swings in the match using the data given and analyze the most related factors.
  - Give advice to a player preparing for a new match against a different opponent considering the variations in past match momentum swings.
- **Problem 4:** Test the developed model on other matches to assess its accuracy. If the model performs poorly, identify potential factors that may need to be included in future iterations. Additionally, evaluate the model's generalizability to different match types and even sports types.
- **Problem 5:** Generate a report containing the discoveries within a concise 25-page limit. Additionally, provide a one- to two-page memorandum summarizing the outcomes, offering guidance to coaches on the significance of "momentum" in tennis and strategies to equip players in effectively responding to disruptions that influence the game's rhythm.

### 1.3 Our Work

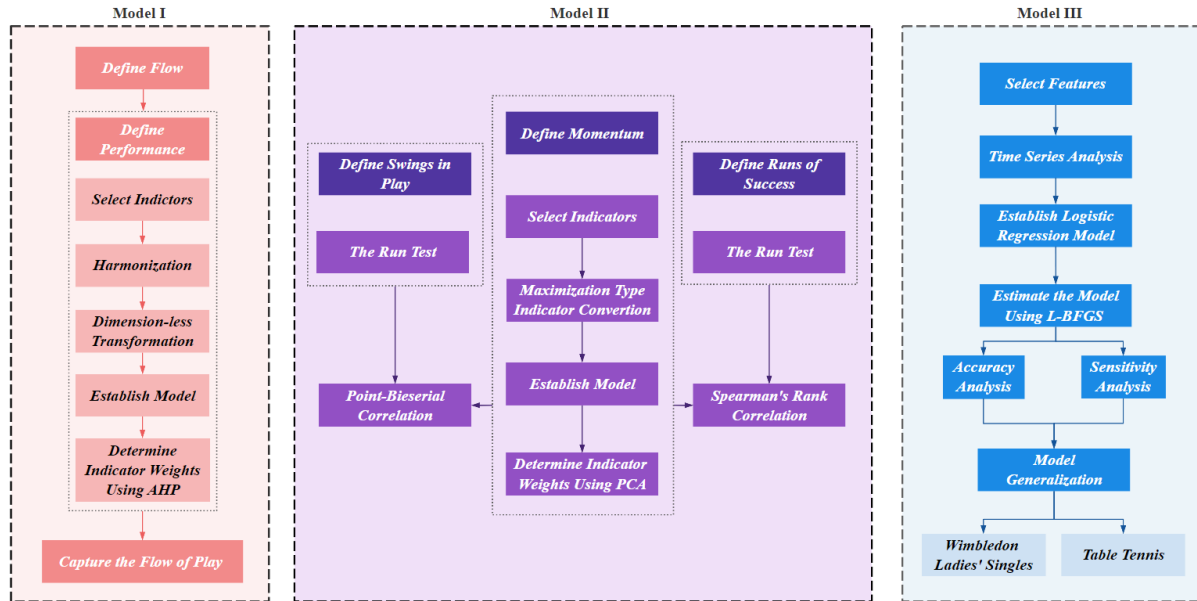


Figure 2: Our Work

## 2 Assumptions and Justifications

**Assumption 1:** Players do not get injured during the match and are not affected by pre-existing injuries.

- Justification: To ensure that the performance and momentum of the players during the match are not affected by any physical limitations or injuries so as to focus on analyzing other factors that contribute to the momentum on the court.

**Assumption 2:** Players are not influenced by external factors during the match.

- Justification: To isolate the impact of external variables during the match and study the intrinsic abilities and strategies of the players so that we can analyze their momentum based solely on their skills and decision-making on the court.

**Assumption 3:** The quality of sports equipment, including tennis rackets and shoes, is comparable between the two players during the match.

- Justification: To eliminate any potential advantage or disadvantage that could arise from equipment variations, which allows us to focus on analyzing the players' abilities and tactics without the confounding factor of equipment quality.

**Assumption 4:** In tennis matches, the serve success rate of players before gaining the right to serve is set at 100%.

- Justification: The serve success rate metric is posterior in subsequent analyses; hence, the initial value is set to the highest possible.

**Assumption 5:** All numerical data related to matches, excluding missing values and outliers, are comprehensively and accurately recorded, with no artificial measurement errors.

### 3 Notations and Glossary

#### 3.1 Notations

Table 1: Notations

Symbols	Definition
$f$	The flow of the game at the point
$F$	The flow of play
$PM$	The performance of player
$N_i$	Judgement matrix.
$m_{ij}$	The momentum of Player $i$ at the $j$ th point
$M_j$	The momentum of the match at the $j$ th point
$G$	Number of games won
$S$	Number of sets won
$P$	Scoring indicator
$E$	Unforced error count
$SW$	Swings in play
$Q$	The number of consecutive scores won
$Q'$	Whether consecutive wins occur

#### 3.2 Glossary

After comprehending the glossary of tennis terms, we summarized the rules of tennis matches in the flowchart as shown in the Figure 3.

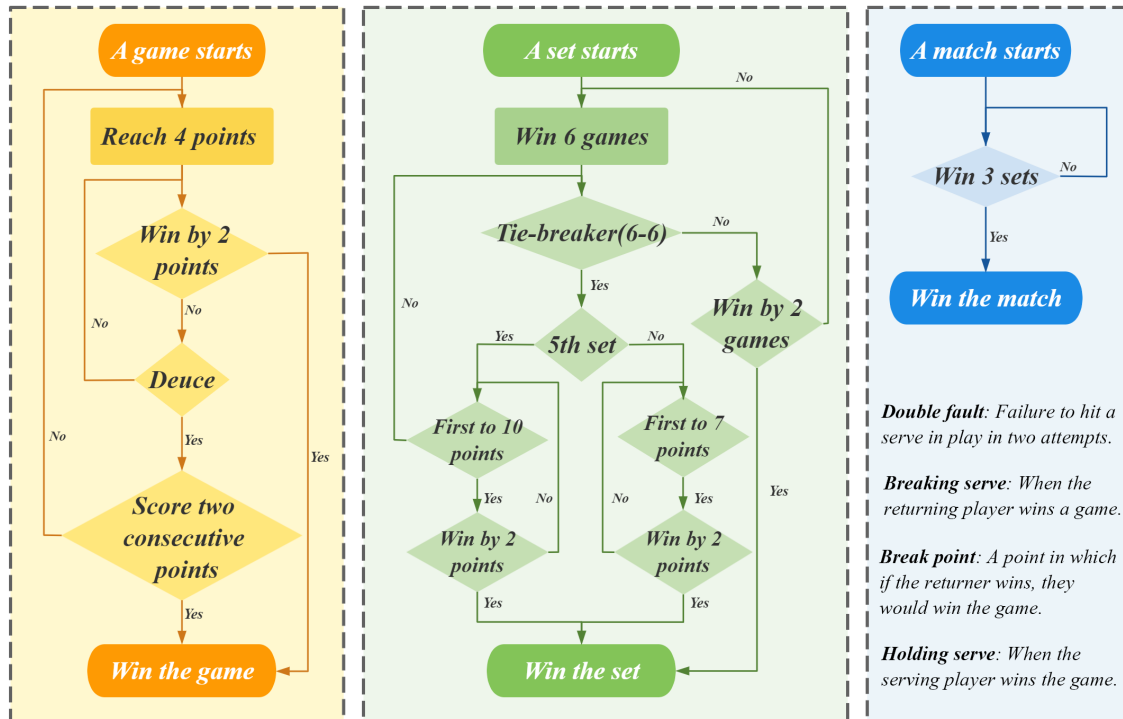


Figure 3: Glossary

## 4 Data Preprocessing

By reviewing the provided data, we found that there are format errors, an outlier and missing values in the table `Wimbledon_featured_matches.csv`

- **Format Errors:** There are format errors in the column named *elapsed\_time* from lines 583 to 636 in the table. These entries incorrectly use a 24-hour format that exceeds the typical hour range, such as "24:56:34", which does not conform to the standard time format of %H:%M:%S. In order to correct these errors, we adjust the time format to a valid range and convert the time data into seconds.
- **Outlier:** There are only two sets played in the match numbered 2023-wimbledon-1403. After searching for the details of this match, we found that one of the players pulled out due to injury. We consider this as an outlier and delete the data of this match.
- **Missing Values:** There are considerable missing values in the column named *winner\_shot\_type*, *peed\_mph*, *serve\_width*, *serve\_depth*, *return\_depth*.

Since rule of traditional tennis scoring varies from that in a tie-breaker, we unified the scoring rules into the tie-break scoring rule for the convenience of the following analysis. We altered the values in the column named *p1\_score* and *p2\_score* into number of points won by Player 1 in a game using Python.

Some examples of the data preprocessing are shown as Table 2.

Table 2: Examples of Data Preprocessing

Line Number	Variable	Initial Value	Correct Value
19	<i>p2_score</i>	AD	4
59	<i>p1_score</i>	40	3
586	<i>elapsed_time</i>	24:56:34	0:56:34
5496	<i>return_depth</i>	NA	/

## 5 Model I: Quantitative Model of Flow and Player Performance Based on AHP to Determine the Indicator Weights

### 5.1 Definition of Flow

Kido<sup>[1]</sup> defined volleyball "flow" as the process of evaluating the consistency between predicting success and its actual outcome. The study identified interconnected concepts like "invisible negative spiral," "invisible positive spiral," "enforceability of referee," and "timewise grace" within the flow structure. Asai and Sagawa<sup>[2]</sup> surveyed experienced volleyball players to explore the factors influencing flow. Their findings revealed that flow is influenced by scores, overall game situations, and contextual factors. Yonezawa and Tawara<sup>[3]</sup> examined how consecutive points resulting from mistakes affect the flow in volleyball, emphasizing the dynamic nature of flow during gameplay.

**We ultimately interpret "flow" as the overall game situation, assuming it is related to the performance of both players.** Here, we provide the definition expression for the "flow" at each point:

$$f_j = PM_{1j} - PM_{2j} \quad (1)$$

where  $f_j$  represents the flow of the game at the  $j$ th point, and  $PM_{ij}$  represents the performance of Player  $i$  at the  $j$ th point ( $i = 1$  for Player 1,  $i = 2$  for Player 2).

The definition expression for the flow of the entire match is as follows:

$$F = \sum_{j=1}^k PM_{1j} - \sum_{j=1}^k PM_{2j}, \quad (2)$$

where  $F$  represents the flow of the entire match.

Afterward, the challenge lies in identifying meaningful indicators to quantify the performance of Player 1 and Player 2.

## 5.2 Definition of Performance

### 5.2.1 Indicators Selection

Performance ( $PM$ ) is a result-oriented concept. Only the performance at a specific scoring point is considered here in the definition, so the performance of an individual player at that certain point may have negative values. Following the principles of **scientificity, systematicity, independence, hierarchy, practicality, and simplicity**, we conducted further analysis on the indicators in the original data dictionary. We select the indicators with strong correlation and finally considered the following aspects of indicators:

- Qualitative Indicators
  - Serving-related Indicators

Indicators	Description	Possible Values
<i>Server</i>	Whether the player has control of the serve, which affects scoring opportunities.	0: False; 1: True
<i>ServerNum</i>	Number of successful serves, which represents the serving proficiency of the player.	1: Once; 2: Twice
<i>Server · ServerNum</i>	As the above two indicators are highly correlated, the product of them will be treated as a single entity in the model for analysis.	0, 1, 2

- Match-related Indicators

Indicators	Description	Possible Values
<i>PointVictor</i>	A combination of <b>scoring-situation-related factors</b> , representing the overall success when scoring a point.	
<i>PointFail</i>	A combination of <b>error-situation-related factors</b> , signifying errors that lead to point losses.	
<i>GameVictor</i>	Whether the player wins the game.	0: False; 1: True
<i>SetVictor</i>	Whether the player wins the set.	0: False; 1: True

- Scoring-related Indicators



Indicators	Description	Possible Values
<i>IFPointVictor</i>	Whether the player wins the point.	0: False; 1: True
<i>PointWinner</i>	Whether the player hits an untouchable winning shot, which indicates an aggressive playing style and proactiveness.	0: False; 1: True
<i>PointAce</i>	Whether the player hits an untouchable winning serve, which is an efficient way to score and often indicates high serving quality.	0: False; 1: True
<i>NetWin</i>	Whether the player scores near the net, which usually requires advanced technique and agility, reflecting the player's skill level more than regular scoring.	0: False; 1: True

◦ Error-situation-related Indicators

Indicators	Description	Possible Values
<i>DoubleFault</i>	Whether the player loses a point due to a double fault, which is a form of self-inflicted loss that negatively impacting the player's mental state and momentum.	0: False; 1: True
<i>PointWinner</i>	Whether the player loses a point due to an unforced error, which may indicate a lack of concentration or poor technical execution by the player.	0: False; 1: True

• Quantitative Indicators

Indicators	Description	Possible Values
<i>RallyCount</i>	The number of rallies, reflecting the intensity and endurance of the game.	Integer number

## 5.2.2 Comprehensive Evaluation Data Processing

• Qualitative Indicators

- Harmonization: Harmonization is to unify the type of evaluation indicators. In qualitative indicators, there exist both maximization indicators (where higher values are better) and minimization indicators (where lower values are better). Due to the negative correlation between *PointFail* and the performance of the player, as well as *Server/ServerNum* and the player's performance, we will use  $(1 - \text{PointFail})$  and  $(2 - \text{Server} \cdot \text{ServerNum})$  in the model to convert minimization variables into maximization ones.

• Quantitative Indicators

- Harmonization: In this case, the *RallyCount* is a maximization indicator, so further processing is not required.
- Dimensionless Transformation: Here, the method of sample transformation will be used.

$$a_{ij}^* = \frac{a_{ij} - \mu_j}{s_{ij}}, 1 \leq i \leq n, 1 \leq j \leq m,$$

$$\mu_j = \frac{1}{n} \sum_{i=1}^n a_{ij}, \quad s_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (a_{ij} - \mu_j)^2},$$

where the sample mean is  $\mu_j$  and the sample standard deviation is  $s_j$ ,  $a_{ij}^*$  is the standardized observation.

### 5.2.3 Mathematical Model Establishment

Eventually, we determined the function expression for the evaluation of player's performance as follows:

$$PM = 100 \times (c_0 + c_1(2 - Server \cdot ServerNum) + c_2 \cdot RallyCount) \cdot (d_1 \cdot PointVictor + d_2 \cdot (1 - PointFail) + d_3 \cdot GameVictor + d_4 \cdot SetVictor) \quad (3)$$

$$PointVictor = a_1 \cdot IFPointVictor + a_2 \cdot PointWinner + a_3 \cdot PointAce + a_4 \cdot NetWin \quad (4)$$

$$PointFail = b_1 \cdot DoubleFault + b_2 \cdot UnforcedError \quad (5)$$

Figure 4 is a visual presentation of the hierarchy of various indicators. We refer to both qualitative and quantitative indicators as variables in the model, and the weights assigned to each indicator as the parameters of the function. Considering that the impact of serves and number of rallies on the model is global, we add the variables mentioned above and incorporate them into the model as a whole. Additionally, we define *PointVictor* and *PointFail* as secondary indicators.

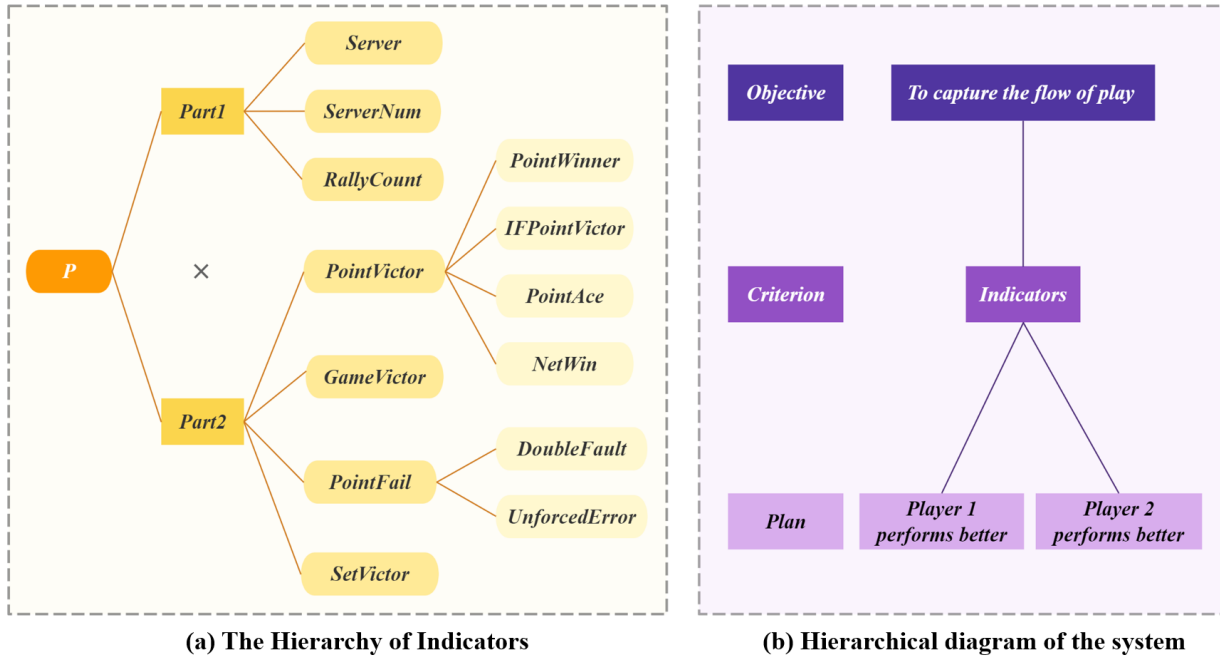


Figure 4: Analytic Hierarchy Process (AHP) diagram

### 5.2.4 Determination of Indicator Weights

We adopt the **Analytic Hierarchy Process (AHP)** to determine the weights of the indicators in the function expression mentioned above. The Analytic Hierarchy Process (AHP) is a decision-making method that, based on a thorough analysis of the nature, influencing factors, and intrinsic relationships of complex decision problems, mathematically formalizes the decision-making thought process using limited quantitative information. It thereby offers a straightforward approach to addressing multi-objective, multi-criteria, or unstructured complex decision problems.<sup>[4]</sup>

### • Determination of Parameters of *PointVictor*

Due to the relationships depicted in the Venn diagram as shown in the Figure 5, namely *IFPointVictor*, *PointWinner*, *PointAce*, and *NetWin*.

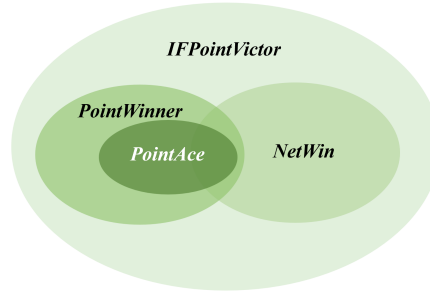


Figure 5: The Relationships among Four Scoring-related Indicators

And considering that the variables in the function expression have an additive relationship, we determine their importance as follows:  $IFPointVictor > PointWinner = NetWin > PointAce$ . The judgement matrix is as follows:

$$\begin{aligned}
 N_1 &= \begin{bmatrix} 1 & 2 & 3 & 2 \\ \frac{1}{2} & 1 & 2 & 1 \\ \frac{1}{3} & \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{2} & 1 & 2 & 1 \end{bmatrix} \xrightarrow{\text{Normalize column vectors}} \begin{bmatrix} 0.429 & 0.444 & 0.375 & 0.444 \\ 0.214 & 0.222 & 0.25 & 0.222 \\ 0.143 & 0.111 & 0.125 & 0.111 \\ 0.214 & 0.222 & 0.25 & 0.222 \end{bmatrix} \\
 &\xrightarrow{\text{Compute the row-wise sum}} \begin{bmatrix} 1.692 \\ 0.909 \\ 0.490 \\ 0.909 \end{bmatrix} \xrightarrow{\text{Normalize}} \begin{bmatrix} 0.423 \\ 0.227 \\ 0.123 \\ 0.227 \end{bmatrix} = W^{(1)} \\
 B_1 W^{(1)} &= \begin{bmatrix} 1 & 2 & 3 & 2 \\ \frac{1}{2} & 1 & 2 & 1 \\ \frac{1}{3} & \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{2} & 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1.699 \\ 0.227 \\ 0.123 \\ 0.227 \end{bmatrix} = \begin{bmatrix} 1.699 \\ 0.911 \\ 0.491 \\ 0.911 \end{bmatrix} \\
 \lambda^{(1)}_{max} &= \frac{1}{4} \left( \frac{1.699}{1.699} + \frac{0.911}{0.227} + \frac{0.491}{0.123} + \frac{0.911}{0.227} \right) = 0.227
 \end{aligned}$$

According to the definition of consistency index,

$$CI = \frac{\lambda^{(1)}_{max} - n}{n - 1}, \quad (6)$$

substituting  $\lambda^{(1)}_{max}$  into the formula yields  $CI_1$  as 0.00346. By querying the Random Index (RI) and calculating  $CR = \frac{CI}{RI}$  with its value, we obtain  $CR = 0.00384 < 0.1$ , indicating that the inconsistency level of  $N_1$  is within the acceptable range. This implies satisfactory consistency, and the consistency test can be passed. Therefore, at the indicator level, the substantial figurative weight of  $B_1$  is represented by  $W^{(1)}$ .

### • Determination of Parameters of $b_1, b_2$

Considering that unforced errors include double faults, we assume that the importance of double faults relative to unforced errors is 2. We determine the judgment matrix as:

$$N_2 = \begin{bmatrix} 1 & 2 \\ \frac{1}{2} & 1 \end{bmatrix}$$

Similarly, we obtain the weight matrix:  $[b_1 \ b_2] = [\frac{2}{3} \ \frac{1}{3}]$ . Since  $n = 2$ , there is no need to perform a consistency test.

• **Determination of Parameters of  $c_0, c_1, c_2$**

We assume that Player 1's importance relative to  $(2 - \text{Server} \cdot \text{ServerNum})$  is 2, relative to *RallyCount* is 3, and the importance of  $(2 - \text{Server} \cdot \text{ServerNum})$  relative to *RallyCount* is 2. The judgement matrix is as follows:

$$N_3 = \begin{bmatrix} 1 & 2 & 3 \\ \frac{1}{2} & 1 & 2 \\ \frac{1}{3} & \frac{1}{2} & 1 \end{bmatrix}$$

We obtain the weight matrix:  $[c_0 \ c_1 \ c_2] = [0.539 \ 0.297 \ 0.164]$

$CR = 0.00794 < 0.1$ , the consistency test can be passed.

• **Determination of Parameters of  $d_0, d_1, d_2, d_3$**

Assuming that the impact of scoring situations (*PointVictor*, *PointFail*) and losing situations is equally important to player performance, and it is higher than the impact of *GameVictor* and *SetVictor*, the judgement matrix is as follows:

$$N_4 = \begin{bmatrix} 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 2 \\ \frac{1}{2} & \frac{1}{2} & 1 & 1 \\ \frac{1}{2} & \frac{1}{2} & 1 & 1 \end{bmatrix}$$

The weight matrix is:  $[d_0 \ d_1 \ d_2 \ d_3] = [\frac{1}{3} \ \frac{1}{3} \ \frac{1}{6} \ \frac{1}{6}]$

$CR < 0.01 < 0.1$ , the consistency test can be passed.

### 5.3 Analysis and Visualization of Results

According to the definition expression of the flow in the entire match in equation (1),  $F > 0$  indicates that Player 1 is more likely to win, and  $F < 0$  indicates that Player 2 is more likely to win. Taking the classic match with "match\_id" as "2023-wimbledon-1701" as an example, we can plot a line graph showing the variation of flow over time in the match, as shown in Figure 6.

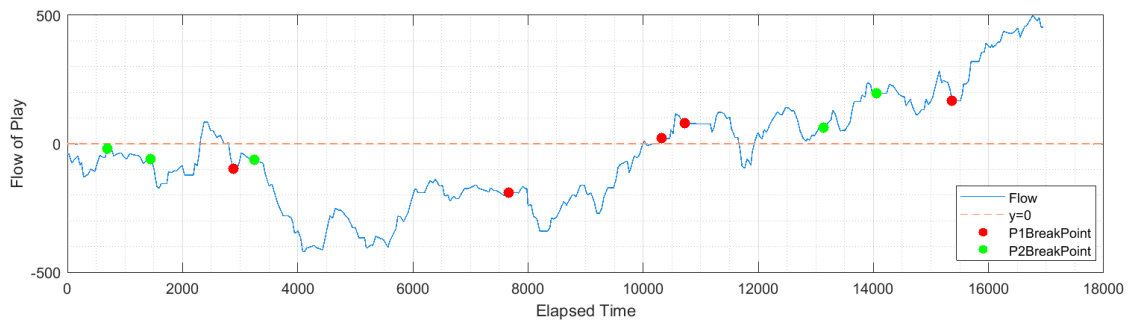


Figure 6: The Flow of Play over Time in Match 1701

The graph includes a reference line at  $y = 0$ . And we can determine which player has the upper hand in the entire match by observing whether the flow is above or below this line. In this particular match, Carlos Alcaraz's comeback after trailing aligns with the actual match situation. We have also marked the breakpoints in the match using green or red dots.

The performance of each player at a given point in the match, which is represented by the flow, can be analyzed using equation (1).  $f_j > 0$  indicates that Player 1's performance is better than Player 2's at the  $j$ th point.  $f_j < 0$  indicates that Player 2's performance is better than Player 1's at the  $j$ th point. The quantification of their respective performances is determined by the specific numerical values of  $PM_{1j}$  and  $PM_{2j}$ . Similarly, Figure 7 visualizes the performance of both players at each point in the match.

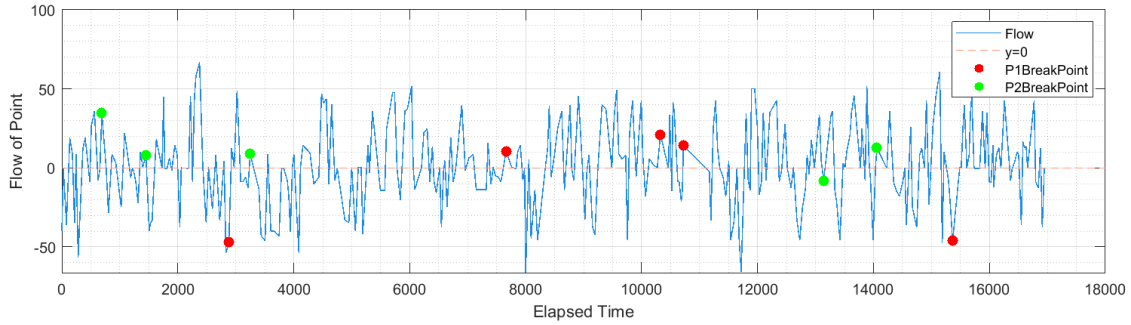


Figure 7: The Flow of Point over Time in Match 1701

## 6 Model II: Momentum Evaluation Model Based on PAC and Statistical Analysis

### 6.1 Definition of Momentum in Tennis Matches

Momentum is used by psychologists to describe a competitive situation between two individuals, which fits perfectly to a sports environment. Iso-Ahola and Mobily, in 1980, offered an alternative definition of psychological momentum. They described it as "an acquired or gained psychological power that alters interpersonal perceptions and influences an individual's mental and physical performance."<sup>[6]</sup> In 1985, Gilovich, Vallone, and Tversky explored the phenomenon known as the "hot hand" in basketball, who specifically associated momentum with a player's shooting accuracy in the context of basketball.<sup>[7]</sup>

We assume that momentum is related to factors such as the number of games won, scoring, and consecutive scoring. Eventually, we define the comprehensive evaluation value of momentum at each point as:

$$M_j = m_{1j} - m_{2j}, \quad (7)$$

where  $M_j$  represents the comprehensive evaluation value of momentum at the  $j$ th point (hereinafter referred to as e-momentum),  $m_{ij}$  represents the comprehensive evaluation value of the momentum of Player  $i$  at the  $j$ th point ( $i = 1$  for Player 1,  $i = 2$  for Player 2). Afterwards, we searched for appropriate indicators to quantify the momentum of Player 1 and Player 2.

### 6.1.1 Indicators Selection

Momentum is the power acquired through movement or a series of events, which accumulates over time. We pick the indicators that impact momentum in the data dictionary and finally considered the following aspects of indicators:

- **Game Indicator  $G$** 
  - Number of games won (0, 1, 2, ...): Reflects the number of games won by the player. An increase in the number of games won is beneficial to a player's confidence and psychological advantage, positively impacting the overall momentum of the match.
- **Set Indicator  $S$** 
  - Number of sets won (0, 1, 2, 3): Reflects the number of sets won by the player.
- **Scoring Indicator  $P$** 
  - Scored points  $p$  (0, 1):  $p$  measures the player's scoring performance in the match. When a player scores consecutively, he/she tends to maintain control of the match and exerts pressure on their opponent. A growing number of scores boosts a player's morale and impetus, influencing their momentum.
  - Consecutive scoring count  $c$  (0, 1, 2, ...):  $c$  represents the number of consecutive scoring instances by the player. Continuously scoring contributes to a player's confidence and strengthens their momentum, enhancing their competitiveness in the match.
- **Error Indicator  $E$** 
  - Unforced error count (0, 1):  $E$  measures the player's unforced errors in the match. Committing unforced errors can lead to a loss of impetus, confidence, weakening a player's overall momentum.

### 6.1.2 Comprehensive Evaluation Data Processing

- Qualitative indicators
  - There are both maximization and minimization types in the qualitative indicators. Since the error indicator  $E$  is negatively correlated with momentum, while other indicators show positive correlations, we use  $-E_j$  in the model to convert the minimization type indicator into a maximization type for processing.
- Quantitative indicators
  - Non-dimensional processing is specific to each individual player, therefore, after processing, it will impact the comparison and judgment of momentum between players. Therefore, non-dimensionalization will not be performed in this case.

### 6.1.3 Function Model Establishment

The function expression determining the momentum of players is as follows:

$$z_{ij} = \alpha_1 \cdot G_{ij} + \alpha_2 \cdot S_{ij} + \alpha_3 \cdot P_{ij} - \alpha_4 \cdot E_{ij} \quad (8)$$

$$P_{ij} = \sum_{k=1}^j p_{ik} \cdot (1 + 0.2c_{ik}) \quad (9)$$

where  $z_{ij}$  is the momentum of the player,  $G_{ij}$  is the accumulated number of games won,  $S_{ij}$  is the accumulated number of sets won,  $p_{ik}$  is the indicator for scoring at the  $k$ th point,  $c_{ik}$  is the consecutive scoring count at the  $k$ th point, and  $E_{ij}$  is the accumulated count of unforced errors. The value of 0.2 is based on a general understanding or empirical observation of the impact of consecutive win streaks on player performance, and it is adjusted based on specific circumstances and needs.

### 6.1.4 Determination of Indicator Weights

The function parameters of the expression were determined using **Principal Component Analysis (PCA)**. Principal component analysis is a versatile statistical method for reducing a cases-by-variables data table to its essential features, called principal components. Principal components are a few linear combinations of the original variables that maximally explain the variance of all the variables.<sup>[5]</sup> The principle of principal component analysis is to find  $\alpha_1, \alpha_2, \alpha_3$ , and  $\alpha_4$  such that  $\text{var}(z_{ij})$  is maximized while satisfying  $\alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \alpha_4^2 = 1$ . The detailed process is as follows:

#### Step 1: Dimensionless Processing using the Sample Transformation Method:

Centralize the variables  $G_{ij}, P_{ij}, S_{ij}, E_{ij}$  to obtain the centralized data matrix  $M_c$ .

#### Step 2: Calculate the correlation coefficient matrix among variables:

Construct the correlation coefficient matrix for the centralized data:  $S_c = \begin{bmatrix} 1 & \rho_{GP} & \rho_{GS} & \rho_{GE} \\ \rho_{PG} & 1 & \rho_{PS} & \rho_{PE} \\ \rho_{SG} & \rho_{SP} & 1 & \rho_{SE} \\ \rho_{EG} & \rho_{EP} & \rho_{ES} & 1 \end{bmatrix}$

Use the general formula for  $\rho_{XY}$ :

$$\rho_{XY} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n - 1}$$

where  $n$  is the number of observations. Plot the heatmap of the correlation coefficient matrix between Player 1 and Player 2, as shown in Figure 8.

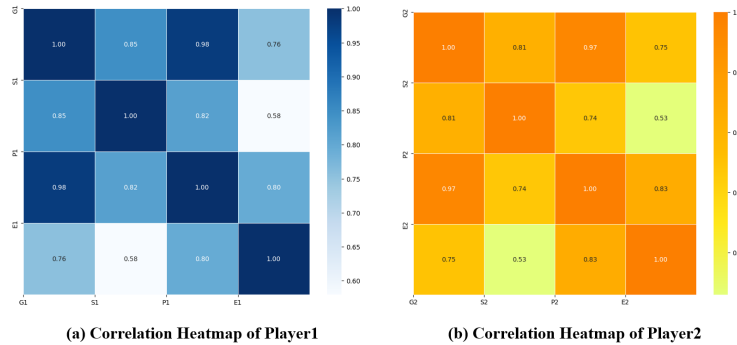


Figure 8: Correlation Heatmap of the Various

#### Step 3: Eigenvalue Decomposition:

Perform eigenvalue decomposition on the correlation matrix  $S_c$  to obtain eigenvalues  $\lambda$  and the corresponding eigenvector matrix  $V$ :

$$S_c V = \lambda V,$$

where  $\lambda$  represents the eigenvalues, and  $V$  is the matrix of eigenvectors.

#### Step 4: Construction of the Principal Component Comprehensive Evaluation Model

After determining that the contribution ratio of the first principal component exceeds 85%, and the contribution ratios of the subsequent three components are significantly smaller, we choose to use only the first principal component to build the evaluation model:

$$m_{ij} = \text{contribution}_{ij} \cdot \vec{v}_{ij} = W[G_{ij}, P_{ij}, S_{ij}, E_{ij}]^T,$$

where  $\vec{v}_{ij}$  is the standardized eigenvector of the first principal component and  $contribution_{ij}$  is its contribution. Since only one principal component is selected,  $W$  happens to be the same as  $\vec{v}$ .

### Step 5: The determination of the weight matrix

We ultimately determine the weight matrix of the model as:

$$W = \begin{bmatrix} \frac{(|\alpha_{11}|+|\alpha_{21}|)}{2} & \frac{(|\alpha_{12}|+|\alpha_{22}|)}{2} & \frac{(|\alpha_{11}|+|\alpha_{21}|)}{2} & \frac{(|\alpha_{14}|+|\alpha_{24}|)}{2} \end{bmatrix} = \begin{bmatrix} 0.531 & 0.114 & 0.695 & 0.470 \end{bmatrix}$$

### 6.1.5 Analysis and Visualization of Results

$M_j = m_{1j} - m_{2j}$  reflects which player has stronger momentum at a given point. If  $M_j > 0$ , it indicates that Player 1 has stronger momentum at point  $j$ , while  $M_j < 0$  indicates that Player 2 has stronger momentum. Taking the final match as an example (match\_id="2023-wimbledon-1701"), the line graph illustrating the changes in momentum over time during the match is shown in Figure 9.

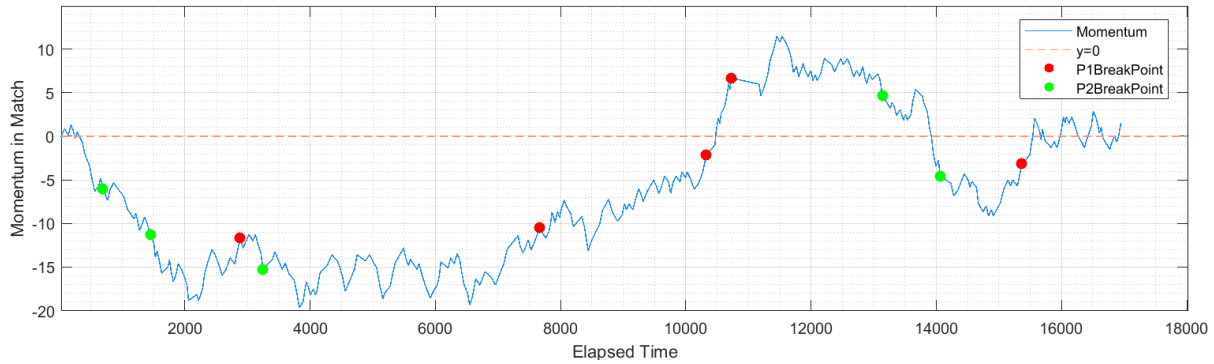


Figure 9: The Momentum in Match over Time

## 6.2 Is there a correlation between swings in play and runs of success with momentum?

Since swings in play, runs of success and momentum theoretically have bidirectional impacts on each other rather than a unidirectional causal relationship, we adopt correlation analysis instead of regression analysis.

### 6.2.1 Swings in Play

Swings in play can be interpreted as changes in the overall trend of the entire match, where a swing refers to the positive or negative change of  $F$  defined in Problem 1. We define it as a binary variable( $SW$ ), representing the occurrence of a swing for each point.

$$SW = \begin{cases} 0, & \text{no swing} \\ 1, & \text{a swing} \end{cases} \quad (10)$$

"Swings in play" refers to the overall dynamics of the entire match. Hence, we analyze its correlation with the momentum of the game( $M$ ).

Since  $M$  is a continuous variable and  $SW$  is a binary variable, we use **Point-Biserial Correlation** to measure their correlation. Point-Biserial Correlation is a method for assessing the



relationship between a binary variable and a continuous variable. Its calculation is similar to Pearson's correlation coefficient.

$$r_{pb} = \frac{\bar{X}_p - \bar{X}_q}{S_t} \cdot \sqrt{pq}, \quad (11)$$

where  $p$  and  $q$  represent the proportions of the binary variable, with  $p + q = 1$ ,  $S_t$  is the standard deviation of the continuous variable,  $\bar{X}_p$  and  $\bar{X}_q$  refer to the mean of the continuous variable for the respective categories of the binary variable, as well as the overall mean.

### 6.2.2 Runs of Success

"Runs of success" can be interpreted as a series of successful events, specifically consecutive scoring. We define a discrete variable ( $Q$ ) to represent the number of consecutive scores (0, 1, 2, 3, etc.). "Runs of success" by one player are analyzed in relation to the player's momentum ( $m$ ). Since  $Q$  is a discrete variable and the player's momentum  $m$  is a continuous variable, we use **Spearman's Rank Correlation** to calculate their correlation.

Spearman's rank correlation coefficient, also known as Spearman's  $\rho$ , is a non-parametric (not dependent on data distribution) statistical measure that quantifies the association between two variables. The calculation formula is as follows:

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)} \quad (12)$$

where  $d_i$  represents the differences in ranks between the two variables, and  $n$  is the number of data points.

### 6.2.3 Correlation Results Analysis

We analyzed the data from 30 matches after data cleaning and calculated the correlation between the occurrence of swings for each point ( $SW$ ) and momentum of the match ( $M$ ), resulting in matrix  $R1$ , which contains **29 correlation coefficients** (as one matches don't have any swings therefore not included in the calculation). We also calculated the significance of these coefficients. In **19 out of 29** the matches,  $SW$  and  $M$  showed a significant relationship.

Similarly, we calculated the correlation between the occurrence of consecutive wins for each point and momentum, resulting in matrix  $R2$ , which contains 30 correlation coefficients. Also, we calculated the significance of these coefficients. Player 1 has significant correlations in **10 out of 30 matches**, while Player 2 has significant correlations in **14 out of their matches**.

The scatter plot in Figure 10 presents three sets of correlation coefficients calculated in the previous context. The significance of each correlation is marked with different colored data points.

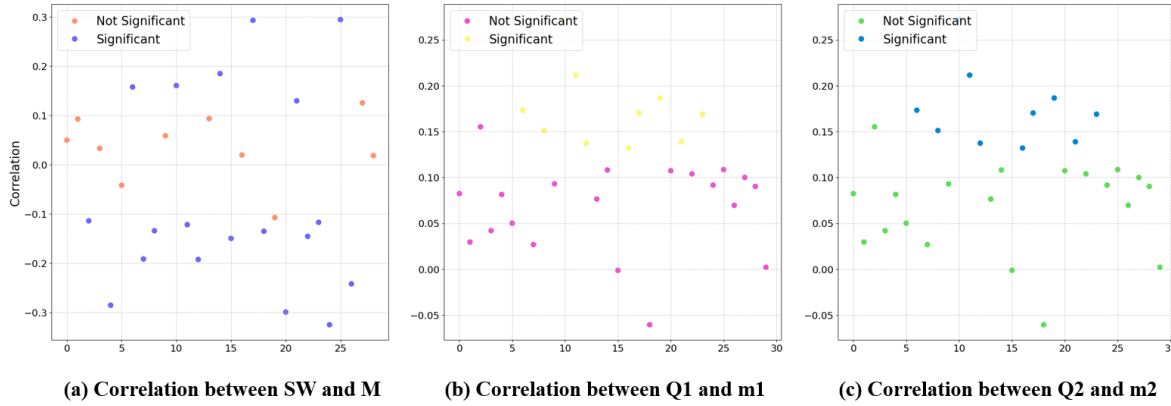


Figure 10: Scatter plot of correlation coefficients

### 6.3 Whether "Swings in Play" and "Runs of Success" are Random

The run test is a method used to test the statistical significance of non-randomness (deviation from randomness) in a binary sequence. Since  $SW$  (swings in play) is a binary variable, we can directly use **the run test**. For the  $Q$  sequence, which is a discrete variable, we consider whether consecutive wins occur when testing randomness, and we transform it into a binary variable  $Q'$ .

$$Q' = \begin{cases} 0, & \text{no consecutive wins} \\ 1, & \text{consecutive wins} \end{cases} \quad (13)$$

The run test for  $SW$  and  $Q'$  is conducted in the following steps:

#### Step 1: Formulate hypotheses

- Null hypothesis ( $H_0$ ): The binary sequence is random and there is no statistically significant non-randomness.
- Alternative hypothesis ( $H_1$ ): The binary sequence exhibits statistically significant non-randomness.

#### Step 2: Calculate the number of runs

A run refers to a consecutive series of the same digits (0 or 1) in the binary sequence. Calculate the number of runs for 0s as  $m$ , the number of runs for 1s as  $n$ , and the total number of runs as  $R$ .

#### Step 3: Calculate the test statistic

Since the number of points in a single match satisfies the conditions for a large sample, calculate the test statistic using the normal approximation:

$$Z = \frac{R - (\frac{2mn}{m+n} + 1)}{\sqrt{\frac{2mn(2mn-m-n)}{(m+n)^2(m+n-1)}}} \sim N(0, 1) \quad (14)$$

#### Step 4: Refer to the distribution table of Z-values to make statistical inferences

Using the above method, we calculated the significance for the data from the matches. The statistical inference results show that out of the 29 matches, the swings of **17 matches** exhibit non-randomness in  $S$ . However, **all of the matches** for Player 1 and Player 2 in the 30-game dataset show the non-randomness of run of success.

## 6.4 Assess the Claim of the Coach

Based on the statistical analysis conducted above, it was found that  $SW$  and  $Q$  were non-random sequences in the majority of the games, suggesting that they were influenced by some factors. However, the correlation analysis between  $SW$  and  $M$ , as well as between  $Q$  and  $m$ , did not reach a significant level, providing no direct evidence of momentum playing a role in the games. Nevertheless, we further conducted a calculation of the Pearson correlation coefficient using the following formula to analyze the relationship between momentum  $M$  and the overall competition trend represented by flow ( $F$ ). It is found that there was a significant correlation in **29 matches**, revealing a certain association between momentum and the game's direction.

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \cdot \sum (Y_i - \bar{Y})^2}}$$

where  $X_i$  and  $Y_i$  are the individual data points of two variables (e.g.,  $M$  and  $F$ ), while  $\bar{X}$  and  $\bar{Y}$  represent their respective means.

So why is the correlation between momentum and swings in play, as well as runs of success, not significant? We speculate that this may be caused by a lag effect in the impact of momentum on the overall game trend ( $SW$ ) and consecutive winning scores ( $Q$ ). It's possible that only when the effect of momentum accumulates to a certain threshold, will it result in changes in  $S$  and  $Q$ .

In conclusion, these results suggest that momentum does play a certain role in the games, but its influence may have a lagging effect. Further research can explore how momentum accumulates and alters the game's direction, while taking other possible factors into account, to gain a more comprehensive understanding of its role in competitive sports.

## 7 Model III: The Swings Prediction Based on Logistic Regression Model Combined with Time Series Analysis

### 7.1 Model Establishment

The variable  $SW$  (whether there is a swing) that needs to be predicted is a binary variable. Therefore, we chose the **logistic regression model**.

Logistic regression is a statistical learning method used to solve binary classification problems. While it's named "regression", logistic regression is in fact a classification algorithm primarily used for predicting binary output variables.

■ **Target Variable:** Whether a swing occur at the point (1) or not (0).

■ **Feature Engineering:** Through the exploration of Problem 1 and Problem 2, we identified the potential factors for swings in the match as the input features for the model, including the following indicators:

- Difference in scores at current point ( $Pdifferent$ )
- Difference in the cumulative average serve success rates at the current point ( $Sdifferent$ )
- Difference in the cumulative walking distances of both players at the current point ( $Ddifferent$ )
- Momentum of the match at the current point ( $M$ )

- **Time Series Component:** Incorporate time series analysis by segmenting the match data into temporal windows (e.g., per game or set) and analyzing trends within these segments.

All the indicators above have an accumulative nature, which represent the influence of previous matches on the current match. We need to examine how the features evolves over the course of the match.

■ **Temporal Features:** Since we are integrating time series analysis, generate time-based features (*pointnumber*).

Incorporate time series analysis by segmenting the match data into temporal windows (e.g., per game or set) and analyzing trends within these segments. **This approach enables logistic regression to predict the outcome of each point based on previous match conditions given by time series analysis.**

## 7.2 Basic Principles

1. **Sigmoid Function:** Logistic regression uses the Sigmoid function (also known as the Logistic function) to map the linear combination of features to the range [0, 1]. The Sigmoid function is defined as:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

where  $z$  represents the linear combination of features.

2. **Linear Combination:** Logistic regression combines the weighted sum of features with the Sigmoid function to obtain the predicted probability:

$$P(Y = 1|X) = \sigma(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)$$

where  $Y$  is the binary output variable,  $X_1, X_2, \dots, X_n$  are the features, and  $\beta_0, \beta_1, \dots, \beta_n$  are the model parameters.

3. **Decision Boundary:** Logistic regression uses a threshold (typically 0.5) to classify positive and negative classes based on the predicted probability.

## 7.3 Model Estimation

In this study, we employed the **L-BFGS (Limited-memory Broyden-Fletcher-Goldfarb-Shanno)** algorithm to estimate the model parameters. L-BFGS is a quasi-Newton method that approximates the Newton method by constructing a second-order approximation of the objective function while avoiding the computation of the full Hessian matrix, resulting in higher computational efficiency in certain cases. We used **80% of the data** as the training set and the remaining **20% as the prediction set**. Matches that  $SW = 1/0$  isn't included in the training set have been excluded (up to 3 matches), predicting whether swing will occur at next point (i.e.,  $SW = 1$ ). We calculated the concordance between the predicted results and the actual results of the last 20% of the data for each match to obtain the model's accuracy:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions of } SW}{\text{Total Number of Predictions}}$$

## 7.4 Results Analysis

In the remaining 27 matches, the model achieved an impressive average accuracy of 97.2%. We simultaneously calculate the error of the model fit. The following formula computes the **Mean Absolute Error (MAE)**:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

where  $n$  represents the number of observations,  $y_i$  is the true value, and  $\hat{y}_i$  is the predicted value. The calculated MAE is 0.0234. We have generated the average feature coefficients of the model, as shown in Table 3:

Table 3: Average Feature Coefficients

Feature	Ddifferent	momentum	Pdifferent	Sdifferent
Coefficient	0.047	0.446	-0.365	0.169

According to the table, it can be observed that momentum has the highest correlation with swings in the match, followed by the impact of score difference on swings in the match.

## 7.5 Advise for a Player Going into a New Match against a Different Player

Considering the variations in "momentum" observed in past matches, we recommend the following strategies for a player when facing different opponents:

1. Emotional control: Based on the experience of momentum changes in previous matches, players should maintain focus and emotional control to face each new match, regardless of the outcomes of previous matches.
  - If the previous match was lost, it is important not to get excessively discouraged. Maintain a positive mindset, follow the usual rhythm of the game, increase energy, and play more actively.
  - If the previous match was won, it is important to think rationally about the victory and adopt wise strategies to avoid overconfidence.
2. Review the previous match and learn to recognize whether the momentum is in their favor, which will help in retaining focus, maintaining emotional control, and staying flexible, as well as adjusting tactics as needed, reacting to the opponent's performance, and utilizing personal strengths to control the pace of the game.

# 8 Model Validation and Generalization

## 8.1 Model Validation

Since the outputs of Model I and Model II are determined by definition, there is no need to analyze their accuracy and sensitivity. Therefore, the following analysis is specific to Model III.

### 8.1.1 Accuracy Analysis

In section 7.4, we have analyzed the accuracy and errors of the logistic regression model (Model III) that we have built. The accuracy of the model was below 0.9 only in the matches with `match_id="2023-wimbledon-1309"` and `match_id="2023-wimbledon-1306"`. We speculate that the poor predictive performance of the model may be influenced by the following factors: **changes in player strategies, variations in players' mental and physical states, and fluctuations in court conditions**. To further understand these influencing factors, we can conduct a detailed analysis of these two matches by watching the live replays of the games and focusing on the aspects where the model failed to make accurate predictions. By comparing the model's predictions with the actual results, we will explore the specific circumstances that could have led to the decrease in accuracy.

In the match with `match_id="2023-wimbledon-1309"`, 26-year-old American player Tommy Paul lost to 22-year-old Czech player Jiri Lehecka. And in the match with `match_id="2023-wimbledon-1306"`, 27-year-old French player Quentin Halys lost to 22-year-old Italian player Jannik Sinner. **Both of these matches involved younger players defeating older players**. We assume that the lower accuracy of the model in these two matches may be influenced by **the age factor of the players**. Evidently, the physical fitness of players tends to decline with age, thus the sensitivity to factors such as distance to run varies greatly among players of different ages. As the matches progress, older players experience faster declines in physical fitness than the younger ones, which is inclined to affect the factors mentioned above, ultimately resulting in lower accuracy in the model's predictions.

In addition, we will also consider **whether the training data of the model adequately covers various match scenarios**. Lastly, there might be special or extreme cases that were not sufficiently included in the model training, resulting in performance degradation in certain matches. For example, Player 1 had an easy victory in his/her previous match and had good physical recovery during the game, while Player 2 had a difficult victory in their previous match and had not fully recovered physically.

### 8.1.2 Sensitivity Analysis

We manually specified a test set ratio of 20%. It's apparent that the change of this value could affect the model training. The sensitivity analysis results are shown in Table 4.

Table 4: Sensitivity Analysis Results

Test set ratio	20%	10%	15%	25%	30%
Average accuracy	96.663%	98.033%	97.834%	97.595%	97.575%

The results indicate that by changing the proportion of the test set, the rate of change in average accuracy is consistently less than 1%. This demonstrates that our model is insensitive to variations within 0.5 times its own range for this parameter, highlighting its robustness and reliability.

## 8.2 Model Generalization

We have extended the application of our model to other competitions. The 2023 Wimbledon Ladies' Singles Championship was one of the most anticipated tournaments, attracting numerous elite players. Czech prodigy Marketa Vondrousova, only 25 years old, became the first unseeded player to reach the Wimbledon Women's Singles final. She faced Tunisia's renowned player, Ons Jabeur, in the final match. Ultimately, Vondrousova defeated Jabeur with a score of 2-0, winning the Wimbledon Ladies' Singles title. For the performance of our model in 30 women's tennis matches (data sourced from the website [www.wimbledon.com](http://www.wimbledon.com)), the average prediction accuracy reached **95.33%**. The predictions for the final set of Marketa Vondrousova were highly consistent with the actual match outcome. Additionally, we evaluate the model's performance in table tennis matches. Since the table tennis match data only includes scoring data (sourced from the website [www.worldtabletennis.com](http://www.worldtabletennis.com)), we have removed the error indicator  $E$  from our model's definition expression. After the modification, the average prediction accuracy of the model is **94.73%**. The calculation results is presented in a violin plot and a scatter plot overlay chart, as shown in Figure 11. This result demonstrates the excellent performance and generalizability of our model across different genders and different types of the matches. Future supplementary analysis directions include evaluating the predictive effectiveness for specific player groups, assessing the influence of different match venues, and comparing the performance of different models in matches in different time periods (e.g., morning, evening). Through these in-depth studies, we can comprehensively assess the performance and applicability of the model.

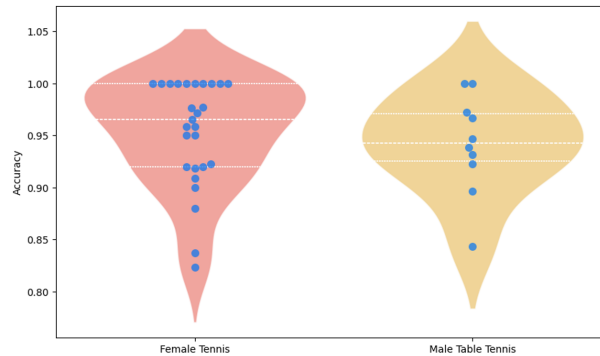


Figure 11: Scatter plot of correlation coefficients

## 9 Model Assessment

### 9.1 Strengths

1. Model I and Model II thoroughly considered the impact of various factors on flow and momentum, determining the weights of each indicator through a rational approach.
2. The established momentum model has a certain predictive effect on swings, aiding in the analysis of game situations and forecasting match outcomes.
3. Exploring the randomness of variables, in addition to conducting tests for the randomness of sequences, correlation tests were performed based on the practical meanings of variables.
4. Model III demonstrates extremely high accuracy in predicting swings during matches, and its robustness and reliability were confirmed through sensitivity tests. Additionally, the model can be extended to other matches, exhibiting applicability across genders and match types.

## 9.2 Weaknesses

1. The excessive number of model indicators increases the difficulty for decision-makers to assess match situations using multiple metrics.

# 10 Memo

**From:** Team 2422350, MCM 2024

**To:** Tennis Coaches

**Date:** February 5, 2024

**Subject:** Understanding the Role of "Momentum" and Preparing Players to Respond to Events Impacting the Flow of Play in Tennis Matches

---

Dear tennis coaches, we are honored to offer suggestions on the role of momentum and how to prepare players to respond to events that impact the flow of a play during a tennis match after data analysis and modeling.

### Suggestion on the Role of Momentum

**We found out in our studies that swings in play and runs of success by one player are not random. There is a certain correlation between momentum and the flow of the game. Additionally, momentum has a delayed impact on the flow of the game and the runs of success.** Only when the cumulative effect of momentum reaches a certain threshold does it change the flow of the match. Moreover, the difference in scores, the difference in serving success rates between the two sides, the difference in cumulative running distance of the players, and the game's momentum are four indicators that can help determine when the flow of play is about to change from favoring one player to the other. Momentum and the difference in scores are most strongly correlated with swings in the match.

By understanding momentum and the role of in the matches, players could employ methods and tactics in game to ensure they're in control of Momentum, rather than a victim of it. We divide momentum into 5 stages with suggestions separately.

1. When momentum is totally against a player. Slowing down, being deliberate and follow rituals
2. When momentum is turning against a player. Increasing energy and playing more aggressively.
3. When momentum is neutral. Both players are vying to not only win points, but also seize control of the momentum. As such, both will tend to play primary patterns on serve and return.
4. When momentum is about to favor the player. Understand how to reach this point. Some players struggle to convert from this stage, perhaps falling into a fake sense of security or over-confidence.
5. When momentum is totally with a player. Opponent is most likely to change tactics. The player on top should also seek to mix up their game to keep their downtrodden foe on the back foot.

At the moment the momentum starts to influence a match, players should begin to recognize which of those five stages they're at, and contextualize their situation, so as to retain focus and maintain control of their emotions

### Suggestions on Preparing Players for Flow-Impacting Events in Tennis Matches

To help players respond effectively to events that impact the flow of play and manage momentum, coaches can implement the following strategies:



## 1. Mental Preparation:

- Mental conditioning: Incorporate mental conditioning exercises into training sessions to help players develop resilience and focus.
- Emphasize mental resilience: Encourage players to develop a positive mindset, focus on the present moment, and maintain confidence even during challenging situations.
- Visualization exercises: Guide players to visualize successful outcomes and mentally rehearse responding to momentum shifts.
- Stress management techniques: Teach players relaxation and breathing exercises to manage stress and keep a clear mind.

## 2. Tactical Approaches:

- Constructive feedback: Provide players with specific feedback on their performance during momentum shifts and suggest adjustments for future matches.
- Analyze opponents' strategies: Help players understand their opponents' rituals and identify potential swing opportunities.
- Adaptability and flexibility: Teach players to adapt their game plan and tactics based on the current momentum and changing circumstances.
- Momentum-building shots: Encourage players to develop shots or tactics that can help regain momentum or disrupt opponents' momentum.

## References

- [1] Kido, T. (2012). The sociological study about "NAGARE" in volleyball game - based on the qualitative analysis about conversation data of college players (in Japanese). *Volleyball Res.*, **14**(1), 28-35.
- [2] Asai, Y., & Sagawa, M. (2013). Process of "streaks" in the volleyball game and game situation (in Japanese). *Coaching Stud.*, **27**(1), 9-22.
- [3] Yonezawa, T., & Tawara, N. (2010). A study on "the game flow" of the volleyball: on the relation between the continuous lost-point and victory or defeat (in Japanese). *Sports Health Sci. Res.*, **41**, 1-7.
- [4] Vargas, L. G. (1990). An overview of the analytic hierarchy process and its applications. *European Journal of Operational Research*, **48**(1), 2-8. DOI: 10.1016/0377-2217(90)90056-H.
- [5] Greenacre, M., Groenen, P.J.F., Hastie, T., et al. (2022). Principal component analysis. *Nat Rev Methods Primers*, **2**, 100. DOI: 10.1038/s43586-022-00184-w.
- [6] Iso-Ahola, S. E., & Mobily, K. (1980). "Psychological momentum": A phenomenon and an empirical (unobtrusive) validation of its influence in a competitive sport tournament. *Psychological Reports*, **46**(2), 391-401.
- [7] Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, **17**(3), 295-314.

## Report on Use of AI

### 1. OpenAI *ChatGPT* (Sept, 2021 version, GPT-3.5 Turbo)

Query: What function in the `scipy.stats` library calculates the Pearson correlation coefficient? Can you provide an example of how to use it?

Output: In the `scipy.stats` library, you can use the `pearsonr` function to calculate the Pearson correlation coefficient. Here's an example of how to use it:

---

```
import scipy.stats as stats

# Two arrays representing two variables
x = [1, 2, 3, 4, 5]
y = [2, 4, 6, 8, 10]

# Calculate the Pearson correlation coefficient and p-value
corr_coeff, p_value = stats.pearsonr(x, y)

# Print the results
print("Pearson correlation coefficient:", corr_coeff)
print("p-value:", p_value)
```

---

```
Pearson correlation coefficient: 1.0
p-value: 0.0
```

---

In this example, we have two variables `x` and `y`, and we want to calculate the Pearson correlation coefficient between them. The `pearsonr` function returns two values: the correlation coefficient and the corresponding p-value. In this case, since `x` and `y` have a perfect positive correlation (i.e., they are perfectly linearly related), the correlation coefficient is 1.0, indicating a strong positive linear relationship. The p-value of 0.0 suggests that the correlation is statistically significant.

### 2. OpenAI *ChatGPT* (Sept, 2021 version, GPT-3.5 Turbo)

Query: How to extract the decimal before the comma in a string? Can you provide me with an example?

Output: To extract the decimal part before the comma in a string, you can use various programming languages or regular expressions. Here's an example using Python:

---

```
import re

def extract_decimal_before_comma(string):
    match = re.search(r'\d+(?=',)', string)
    if match:
        return match.group()
    else:
        return None

# Example usage
string = "The decimal before the comma is 3.14159,"
decimal = extract_decimal_before_comma(string)
print(decimal) # Output: 3.14159
```

---

In this example, we're using the `re.search()` function from the `re` module in Python to search for a pattern in the string. The regular expression `r'+(?=,)'` matches one or more digits (+) that are followed by a comma ((?=,)). The `(?=,)` is a positive lookahead assertion, which means it checks if the matched digits are followed by a comma without including the comma in the match.

If a match is found, the function returns the matched decimal part before the comma. If no match is found (e.g., if there is no comma in the string or no decimal part before the comma), the function returns `None`.

3. OpenAI *ChatGPT* (Sept, 2021 version, GPT-3.5 Turbo)

Query: Can you give me an introduction to the important tennis events?

Output: Certainly! Tennis is a highly popular and globally recognized sport with a rich history. There are several important tennis events that take place throughout the year, showcasing the skills of the world's top players. Here are some of the most significant tennis events:

1. **Grand Slam Tournaments:** The Grand Slam tournaments are considered the most prestigious events in tennis. There are four Grand Slam tournaments, each held once a year: - Australian Open: Held in Melbourne, Australia, in January. - French Open (Roland Garros): Held in Paris, France, in May and June. - Wimbledon: Held in London, England, in June and July. - US Open: Held in New York City, United States, in August and September.

2. **ATP Tour and WTA Tour:** The ATP (Association of Tennis Professionals) Tour and WTA (Women's Tennis Association) Tour consist of a series of tournaments held throughout the year. These tournaments vary in size and ranking points awarded, with the most prestigious events being the Masters 1000 and Premier Mandatory/Premier 5 tournaments. The ATP Tour and WTA Tour culminate in the season-ending ATP Finals and WTA Finals, respectively, where the top-ranked players compete for the title.

3. **Davis Cup:** The Davis Cup is an international men's team competition that takes place annually. Teams from different countries compete in a knockout format to win the prestigious Davis Cup title. The tournament has a storied history and is considered one of the most significant events in men's tennis.

4. **Fed Cup (Billie Jean King Cup):** The Fed Cup, recently renamed the Billie Jean King Cup, is the equivalent of the Davis Cup for women. National teams compete against each other in a knockout format to win the Fed Cup title. It is one of the most prominent team events in women's tennis.

5. **Olympic Games:** Tennis is also a part of the Olympic Games, held every four years. Both men and women compete in singles and doubles events, representing their respective countries. The Olympic tennis tournament offers players a chance to win a coveted gold medal.

These are just a few of the major tennis events. Other notable tournaments include the ATP and WTA Masters 1000 events, the ATP and WTA 500 events, and the Hopman Cup (mixed-gender team event). Each of these events attracts top-ranked players and contributes to the excitement and growth of the sport.