Problem Chosen	2024 MCM/ICM Summary Sheet	Team Control Number
C		2429961

Momentum—the psychological and physical state of an athlete after a period of high performance—is thought to be highly influential in the flow and outcome of tennis matches. However, few agree on a precise definition of momentum, or if momentum exists at all. By exploring data from Wimbledon 2023 men's matches, we can begin to answer key questions about the nature of momentum: Does momentum play a role in a match, or are all supposed "swings" random? Can we find indicators to predict swings? How should tennis players respond to changes in the flow of play? In this paper, we develop a Markov model that uses Bayesian estimation to predict the next score in a game, then pin down precise definitions of "performance" and "momentum" in order to make predictions about when such swings might occur.

The probabilities of transitioning from one score to another in our Markov model are obtained using Bayesian logistic regression. To create the Bayesian logistic regression model, we selected three attributes—player Elo, serve speed, and fatigue—which, based on our review of the literature, we suspected would act as predictors for point outcomes.

The Bayesian approach seamlessly encodes prior knowledge of tennis, like the substantial in-game advantage held by a serving player, and also offers a framework of certainty and uncertainty that is interpretable to our intended audience of athletes and coaches. Notably, generating probability distributions instead of deterministic score predictions allows us to identify key factors that might incrementally, but not insignificantly, improve the chances of winning a point.

Using expected scores from the Bayesian logistic regression model, we then developed a unique method to evaluate player performance at any moment. By taking the difference between the actual score and expected score, then aggregating differences using an exponentially weighted moving average, we quantified how much a player was exceeding or underperforming expectations at any given point in time.

To assess the effect of momentum on player performance, we measured momentum as the degree to which a player altered the likelihood of winning a game. We compared momentum with player performance score over the next five points and found no causal relationship between the values.

Swings in play were characterized as the 15th percentile of the largest changes in performance scores over a span of five points. Using the same logistic regression model but different indicators that we initially hypothesized could cause improvement or declines in play, we found that the indicators lacked any predictive power in forecasting swings of play.

Our Markov model is able to generate reasonable distributions for observed data and make predictions on new data with 69% accuracy. The accuracy, relatively high for sports predictions, can be attributed to the strength of our priors about the importance of serving, highlighting the advantage of Bayesian methods. Based on our definitions of performance and momentum, we found no evidence overall that momentum or indicators of performance swings impact the players' performance in the given dataset.

Keywords: **Markov chain, MCMC, Bayesian inference**

A Bayesian Analysis Approach to Point Dynamics and Momentum in the 2023 Wimbledon Championships - Men's singles

Contents

1 Introduction

Most athletes, coaches, and sports fans are familiar with *momentum*, the phenomenon where one contestant consistently outperforms the other over a period of time. This mysterious force is behind most exceptional or electrifying moments of gameplay, yet its exact definition remains unclear. Momentum may be unrelated to the actual score, fluctuate quickly between points or slowly across games, and could even just be a state of mind. Still, many agree that momentum has psychological and physical effects that determine the trajectory of a game.

Momentum is often observed in tennis, where extended *matches*—each consisting of five *sets*, which have at least six *games*—against the same opponent give players plenty of time to fall in and out of momentum's "favor." A championship tournament like Wimbledon, where tennis players regularly reach new athletic heights, presents an ideal opportunity to empirically investigate the causes of momentum—and whether it exists at all.

In this paper, we develop a model to tracks player performance during the flow of play, and apply this model to real data from the 2023 Wimbledon Championships to determine the effect of momentum on match results. We begin by listing fundamental assumptions of our model and describing our dataset. In Sections 2 and 3, we present the technical details of our model and the "performance score" metric. In Section 4, we formally define momentum and describe its influence on match outcomes. In Section 5, we explore potential indicators for predicting swings in performance, and we conclude by discussing our method's ability to generalize to other matches in Section 6.

1.1 Assumptions

- *All players are in identical physical condition.* In particular, all players will experience the same effects of fatigue and rest. This implies that any factor that affects both players simultaneously, such as if a game occurs earlier or later in a match, will not be a differentiating attribute. Thus, *we can use any game-specific model to model the entire match.*

- *Slower serves have an equal and opposite effect on the outcome of a point as faster serves.* In particular, any deviation from the median serve speed has the same magnitude of effect on the outcome of a point. Thus, *we can use a normal distribution, which is symmetric around the median serve, to describe how much serve speed affects the outcome.*

- *The stakes of the point have no effect on the players' performance.* Specifically, players attempt to play at their best no matter the situation Thus, *we can apply the same logistic regression model to every point of the match.*

- *The weather has no effect on the game.* The same strategies and attributes of the game are just as effective no matter the humidity or any other change in the environment of play. *Similarly, the crowd has no effect on the game.*

- *The serve represents the same advantage for every player.* Specifically, all players have the same expected rate of scoring on their serve and any deviation from that is due to their performance.

1.2 Data

Our dataset includes information about every point from all Wimbledon 2023 men's matches after the second round. The dataset had several missing values for serving speed, which we filled with the median speed of 115 miles per hour.

2 Modeling the flow of play

We model the flow of points in a game by combining a modified **Markov chain** with **Bayesian logistic regression**. Each score is treated as an independent "state," and we predict transitions between scores using a logistic model.

Motivation

Briefly, a Markov chain is a finite sequence of *states* for which there is some probability of *transitioning* from each state to the other [4]. We can understand a tennis game as a fixed set of score states, each with two possible outcomes (that is, two transitions with non-zero probability) corresponding to which player wins the next point. Figure 1 shows the Markov model for initial game scores $\{(0, 0), (15, 0), (0, 15)\}$.

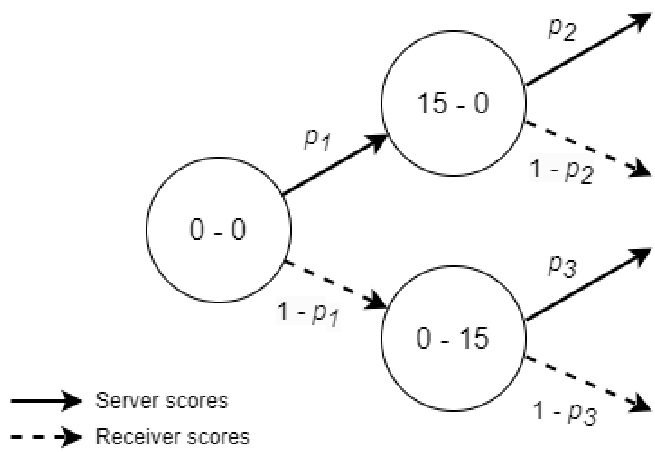


Figure 1: Early gameplay represented as a *Markov chain* between three (`server score`, `receiver score`) states. A given score state i has two possible *transitions*: either the serving player wins the next point with a probability of p_i, or the receiving player wins with a probability of $1 - p_i$.

Why is the Markov model justified? Markov processes are characterized by "memorylessness," meaning predictions about future states can be made entirely based on the current state without any information about the chain's history. Because we assume that players will respond to game conditions in the same way, each score's location in the overall sequence will not have a different effect on each player—and, therefore, no effect on the score's transition probabilities. In practical terms, we use the same regression coefficients to produce probabilities for each state, and differentiate states by choosing predictors that will vary from score-to-score and player-to-player.

2.1 Predictors for point outcomes

Our Markovian assumption implies that to capture the transitions between points, it suffices to consider features of the current state that distinguish the players. All predictor values are normalized to be between -1 and 1.

- **Serving advantage (β_0).** In a tennis match, the server has a higher probability of winning the game. We use this prior understanding of server advantage as a baseline.
- **Elo rating difference (X_E).** The Elo rating is a measure of a player's relative skill level and is often used as a predictor for game outcomes. We obtain the Elo rating for each player prior to the championship from [1], then take the difference between server and receiver Elo to get a single predictor. A positive Elo difference should signal an advantage for the server.
- **Relative serving speed (X_S).** The relative serving speed is the difference between the current serve speed and the median serve speed of the championship, which is 115 miles per hour. This standardization assumes that serves slower than the median have the same absolute effect as faster serves. Intuitively, faster serving speed will make the receiver more likely to lose the point.
- **Fatigue difference (X_F).** The *fatigue* for each player is an exponentially weighted moving average of the previous three distances ran between scores. As with Elo, we take the difference between player fatigues to get a single predictor. Note that a positive fatigue difference means the server runs more than the receiver, so we expect greater differences to correspond with lowered chances of scoring.

2.2 Predicting scores with Bayesian logistic regression

We use Bayesian estimation to generate parameters for our logistic model. For a given score state, let Y be the binary indicator that returns 1 if the serving player wins the next point, and 0 otherwise. This indicator comes from the Bernoulli distribution

$$Y \sim \text{Bernoulli}(p),$$

where p is the probability that the server wins the next point. The logistic model defines the relationship between p and our four chosen predictors by

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_E X_E + \beta_S X_S + \beta_F X_F \tag{1}$$

where $\beta_E, \beta_S, \beta_F$ are the coefficients for the predictors defined earlier. We assume $\beta_0, \beta_E, \beta_S, \beta_F$ are themselves drawn from different normal distributions:

$$\beta_0 \sim \text{Normal}(\mu_0, \sigma_0) \qquad \beta_S \sim \text{Normal}(\mu_S, \sigma_S)$$

$$\beta_E \sim \text{Normal}(\mu_E, \sigma_E) \qquad \beta_F \sim \text{Normal}(\mu_F, \sigma_F).$$

Before fitting the model, we tune the priors by simulating score outcomes under different possible distributions and comparing the simulations to our intuitive beliefs about each predictor (see Section 2.1). Table 1 shows the μ and σ values for each normal distribution after tuning and fitting the model to the observed data.

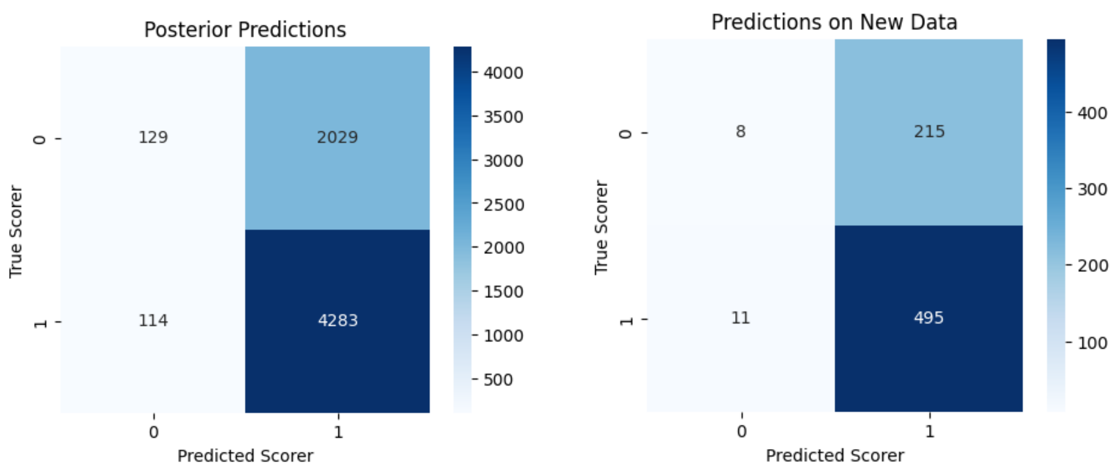
	Prior μ	Posterior μ	Adjusted μ	Prior σ	Posterior σ	Adjusted σ
β_0	0.619	0.787	0.821	0.120	0.027	0.026
β_E	4.464	0.329	0.338	1.000	0.046	0.047
β_S	1.847	1.133	1.243	0.370	0.086	0.094
β_F	-1.115	-0.599	-0.1389	0.250	0.216	0.420

Table 1: Priors and posteriors for the logistic regression coefficients, which are assumed to be drawn from different normal distributions. The "Adjusted" columns gives the posterior values after *doubling* the prior μ during sensitivity analysis.

Errors and uncertainties

To assess the results of our Bayesian logistic regression, we generate a *posterior predictive model* that simulates possible score outcomes by sampling from the parameter posteriors, then compare these simulated values to observed values of Y. Additionally, following best practices in model development, we test our model on unseen data by reserving 10% of our original dataset as a validation set. Figure 2 shows confusion matrices for both results. In general, predictions strongly favor the serving player.

Our model has a 69% accuracy on the validation set. This is similar to our baseline β_0, which represents the pure advantage for the serving player. A more informative measure of uncertainty for probabilistic predictions is the *Brier score*, which is computationally identical to the mean squared error of the predicted probability [3]. Brier scores fall between 0 and 1, and lower Brier scores indicate better model calibration. Our logistic model has a Brier score of 0.208 on the validation set, compared to the baseline with 0.212.



(a) Predictions from the *posterior predictive model*.

(b) Predictions on the validation set.

Figure 2: Confusion matrices for score predictions made by the Bayesian logistic model. In general, the probability of a score is obtained by aggregating predictions from parameters sampled from the posterior. Probabilities greater than or equal to 0.5 will return 1, meaning the serving player wins the next point.

Sensitivity analysis

To examine how changes in the coefficients of priors affect the model's outcomes, we double μ for each prior and assess the changes in distribution. In addition to the "Adjusted" columns of Table 1, results of our sensitivity analysis are shown in Figure 3. We find that the coefficients β_0, β_E, and β_S are robust to prior adjustments, with little change in both mean values and uncertainty. This shows that the data provides stronger evidence for each coefficient than our prior beliefs. On the other hand, the posterior distribution of β_F depends more on our priors, possibly because fatigue is a more complicated transformation of our given data.

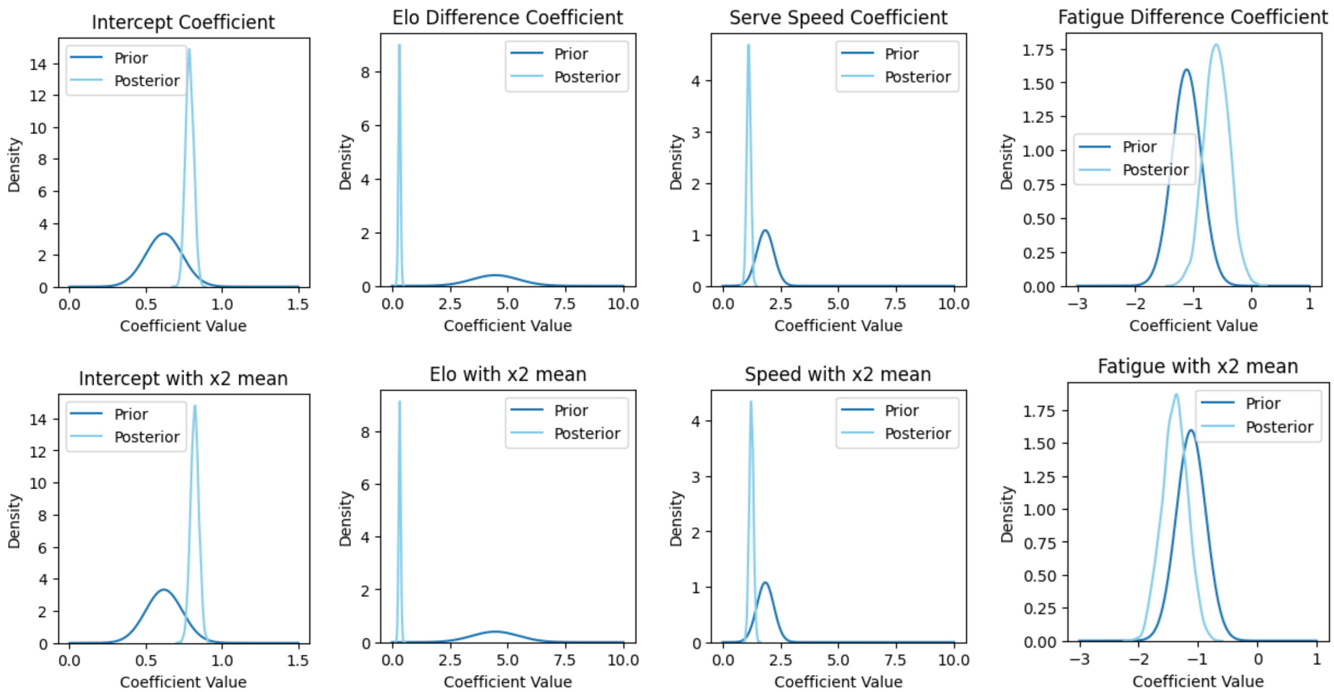


Figure 3: (Top) Prior and posterior distributions for each logistic regression parameter. The posterior distribution represents remaining uncertainty for each parameter after observing real data. (Bottom) Results of sensitivity analysis. Only the fatigue difference coefficient is highly sensitive to changes in the prior.

We can also test the robustness of our model to the assumption at all states are the same, regardless of when in the chain they occur. Table 2 shows the Brier score of different subsets of scores. Each subset represents an in-game situation that could potentially differentiate the score states. For example, when one player has 0 points, they are more likely to be disadvantage and lose the next point. Scores that are only one point apart suggest a more even match, but an early $(15, 15)$ is qualitatively different from a late-game $(\text{AD}, 40)$. The higher subset Brier scores suggest that there are indeed different scores with different relationships to our chosen predictors.

2.3 Visualizing game flow with the Markov model

We now demonstrate our full modeling method by analyzing real match data. For simplicity, we will only apply the model to three (qualitatively different) games, but note that the method offers game-by-game

Subset Description	Brier Score
Validation set with all possible scores	0.208
One player has 0 points	0.348
Scores that are 1 point apart, < 30	0.354
Scores that are 1 point apart, ≥ 30	0.351

Table 2: Analysis of the logistic model's sensitivity to different subsets of possible score states. The *Brier score* is a measurement of accuracy for probabilistic predictions.

insights for an entire match. Figure 4 shows flow charts of each game's Markov chain, along with a graph of the players' performance scores across the entire match. Performance scores quantify how well a player is doing at any point of the match, and will be introduced in the following section.

3 Performance scores

Given a score state i, we want to define a *performance score* P_i that measures the relative performance of the players. In particular, the performance score describes how much each player is exceeding or under-performing expectations. In this section, we explain the steps for computing the final performance score.

3.1 Expected vs. actual score

We first obtain the expected value—that is, the probability of scoring—of the serving player from the logistic model (see Equation 1). We exclude attributes that could inherently influence performance such as Elo, serving speed, and fatigue. Removing these predictors, the expected performance of the server is

$$E_s = \frac{1}{1 + e^{-\beta_0}} = 0.69,$$

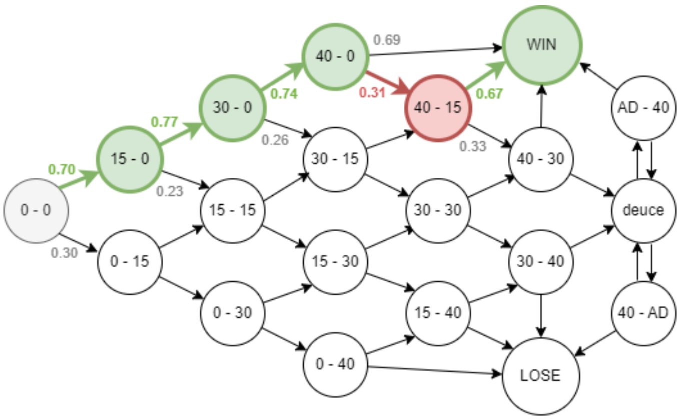
and it follows that the expected performance of the returner is $E_r = 0.31$. To measure how well each player did for that point, we simply subtract the expected score from their actual score. Table 3 shows all possible values of this difference, denoted by O_s and O_r for server and receiver, respectively.

	Server scored: $A_s = 1, A_r = 0$	Returner scored: $A_s = 0, A_r = 1$
O_s	$O_s = A_s - E_s = +0.31$	$O_s = A_s - E_s = -0.69$
O_r	$O_r = A_r - E_r = -0.31$	$O_s = A_r - E_r = +0.69$

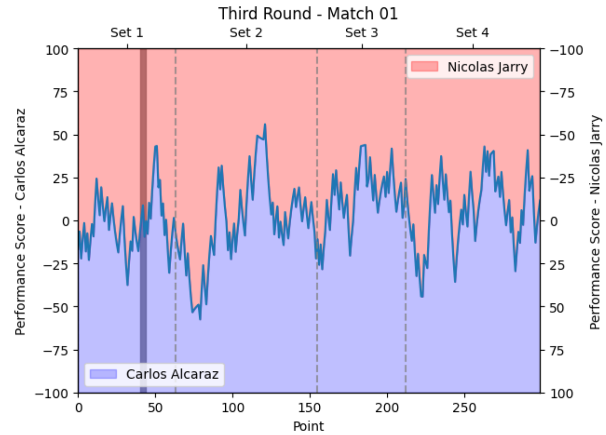
Table 3: All values of O the difference between a player's expected score E and actual score A at a single state.

3.2 Performance as an exponential weighted moving average

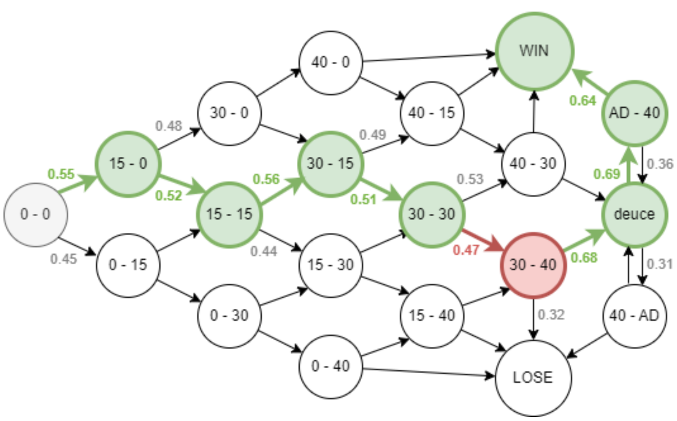
Rather than capturing a player's condition at a single state, which contains little information other than whether they won or lost the point, the performance score should be dynamically updated over time.



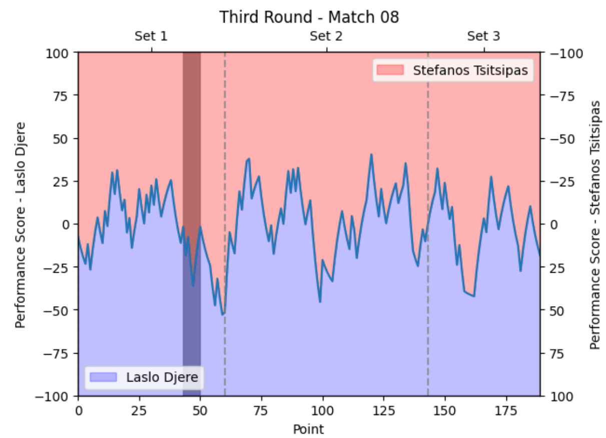
(a) Alcaraz vs. Jarry (third round, match 1, set 1, game 7). Alcaraz is the server and wins the match.



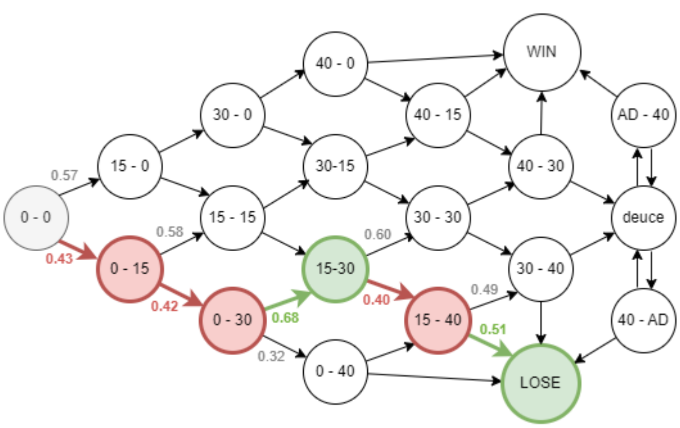
(b) Performance score measurements for Alcaraz vs. Jarry, with game 7 highlighted.



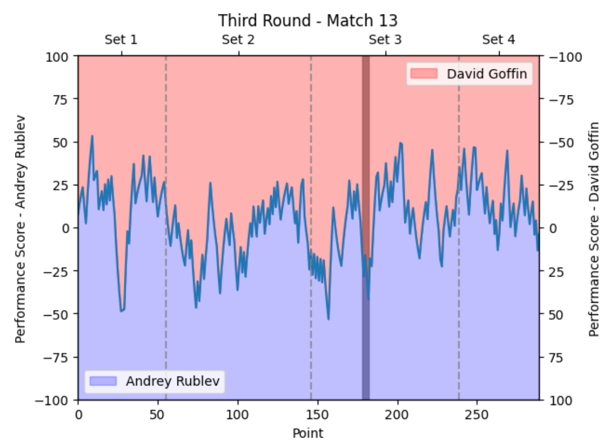
(c) Djere vs. Tsitsipas (third round, match 8, set 1, game 8). Djere is the server and wins the match.



(d) Performance score measurements for Djere vs. Tsitsipas, with game 8 highlighted.



(e) Rublev vs. Goffin (third round, match 13, set 3, game 5). Rublev is the server and loses the match.



(f) Performance score measurements for Rublev vs. Goffin, with game 5 highlighted.

Figure 4: (Left) Markov chain models for individual tennis games. Colored nodes depict the flow of play from real game data, with the associated probabilities listed by each possible transition. *Green* nodes represent correct predictions from our logistic model, and *red* nodes represent incorrect predictions. (Right) Point-by-point changes in *performance score* over an entire match.

We achieve this using an *exponentially weighted moving average* (EWMA) of the difference between the expected score and actual score of each player, which was defined in the previous section.

Now, given a state i, the **performance score** of a player at i is defined by

$$P_i = \alpha \cdot (A_i - E_i) + (1 - \alpha) \cdot P_{i-1}.$$

We set the *smoothing factor* α to 0.16779 so that the most recent five points will never contribute less than 50% to the performance score.

Finally, normalizing performance scores between -100 and 100 allows for a clear and interpretable comparison of player performance. A score of 100 is the theoretical maximum that would be approached if the returner were to score every point, while a score of -100 is the theoretical minimum that would be approached if the server were to lose every point. This nuanced approach to performance measurement paves the way for deeper insight into the point-by-point flow, offering an effective tool for evaluating match results.

4 Momentum

Momentum, in the manner that is used in sports, is a qualitative aspect to describe which team or player is seemingly in control of the game. To determine if this qualitative concept of momentum influences the outcome of games it is essential to first define it quantitatively in a way that captures the essence. In the following sections we define momentum and then explore its impact. Subsequently, we delve into alternative interpretations of momentum to determine if they offer any additional understanding of its effects.

4.1 What is an impactful play?

Momentum is most often associated with big plays. In football, for example, some might say the momentum changes after an interception is thrown or after fumble recovery by the opposing team. To determine what a big play is we first find the probability of winning the game from each state, shown in Table 4, and denoted with the notation S_i.

State (S_i)	**Winning Probability**	**State (S_i)**	**Winning Probability**	**State (S_i)**	**Winning Probability**	**State (S_i)**	**Winning Probability**
$S_1 : (0,0)$	0.5185	$S_6 : (15,15)$	0.4862	$S_{11} : (0,30)$	0.1689	$S_{16} : (40,\text{AD})$	0.1971
$S_2 : (15,0)$	0.7124	$S_7 : (30,15)$	0.7085	$S_{12} : (15,30)$	0.2776	$S_{17} : (\text{AD},40)$	0.7981
$S_3 : (30,0)$	0.8717	$S_8 : (30,30)$	0.4970	$S_{13} : (30,40)$	0.2518	$S_{18} : (40,40)$	0.5011
$S_4 : (40,0)$	0.9863	$S_9 : (40,15)$	0.9160	$S_{14} : (0,40)$	0.0376	S_{19} : Win	1
$S_5 : (0,15)$	0.3029	$S_{10} : (40,30)$	0.7554	$S_{15} : (15,40)$	0.1019	S_{20} : Lose	0

Table 4: The probability of winning the game for a given game states, S_i.

Following this, we analyze how the probability of winning the game shifts with each transition from one state to another, denoted by C_{ij}. This shift in probability signifies the importance of each point, as it quantifies the change in likelihood of winning the game.

$$C_{ij} = S_j - S_i$$

The larger C_{ij} is the more that point increases the odds of the player winning the game and the smaller C_{ij} is the more that point decreases the odds of the player winning the game demonstrating its impact.

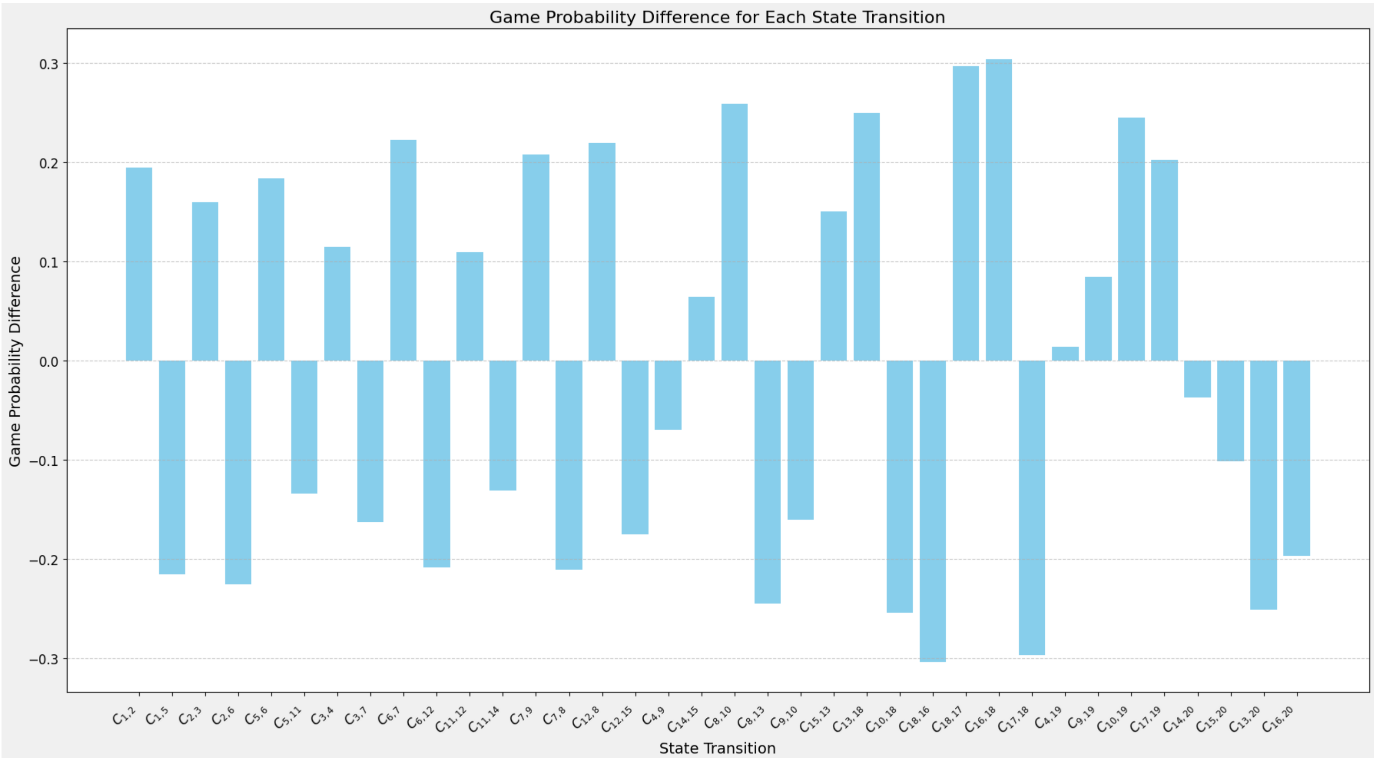


Figure 5: Shows the values of all C_{ij}, which measure
how much a point contributed to the probability of winning a game

4.2 Defining momentum

We then define momentum by how much change in the probability of winning, C_{ij}, has occurred and built up. To do this we use an exponentially weighted moving average (with $\alpha = 0.3$) of C_{ij} of all previous points within the match. This method was chosen so that we accurately capture the *locality* of momentum–a feature that is present in discussions about how momentum could affect outcomes.

$$\mathrm{M}_i = \alpha \cdot (C_i) + (1 - \alpha) \cdot \mathrm{M}_{i-1}$$

4.3 Player performance changes

To assess if momentum affects a player's performance we must look at how their performance changes from a given point in time. This involves comparing the player's performance score at a future point, designated as $t = i + 5$, with their performance score at the initial time, $t = i$. Given that the performance score is calculated as a moving average this approach has the benefit of capturing the player's performance for all those five points, providing a comprehensive assessment of their play level.

$$\Delta P = P_{i+5} - P_i$$

4.4 Analysis

In Figure 6, we present a graphical representation of M in relation to changes in their performance ΔP. This analysis seeks to understand the potential influence of a player's momentum on their subsequent performance levels. Both the visual interpretation of the graph and the application of linear regression reveal a negative correlation, accompanied by considerable variability (indicated by the low r^2 value), between M and ΔP. It's important to note that this correlation should not be construed as causal evidence that momentum adversely affects performance. Rather, this trend may reflect a regression to the mean. The fact that in order to build up a high M value you must be playing exceptionally well means that the decrease in ΔP may just be attributed to the player regressing to their mean. If this were not the case, an endless improvement in performance could ensue, potentially leading to a perpetual feedback loop of increasing performance.

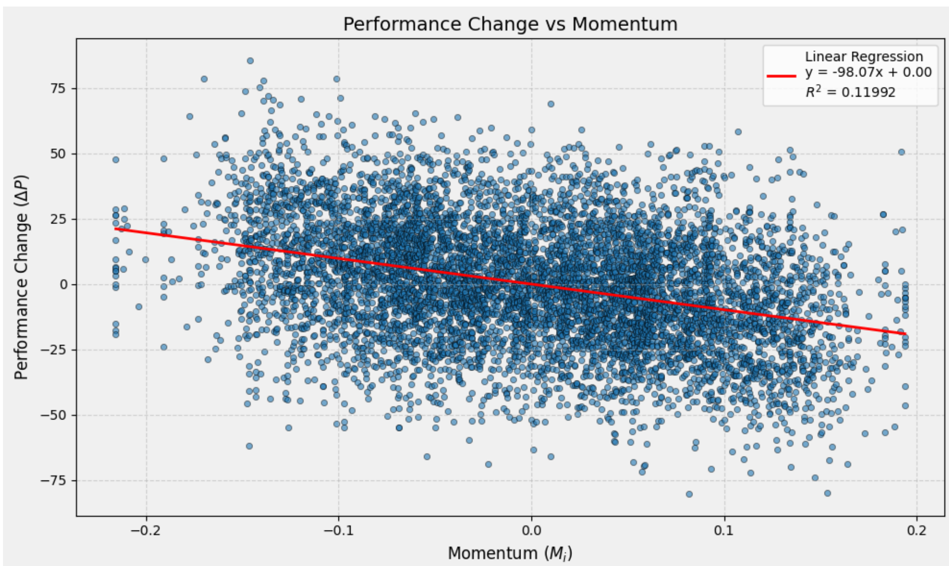


Figure 6: Shows the relationship between Performance Change and Momentum as defined in section 4.3 and 4.2, respectively

4.5 Is momentum something else?

Maybe our definition for momentum, M, was not actually capturing the phenomenon the people talk about, maybe it isn't something that builds up but instead is a phenomenon that is only observable at discrete times. To explore this, we employ the metric C_{ij}, previously defined to quantify the impact of individual plays. By comparing impact plays versus the change in performance score we can see if scoring or not scoring the more important points will cause a player to play better or worse. Figure 7 illustrates a discernible negative correlation between C_{ij} and the change in performance score (ΔP). This correlation, however, can be attributed to the inherent nature of the performance score (P), which partially reflects the outcomes of preceding plays. Specifically, a positive impact play correlates with an immediate positive alteration in the performance score. so, when examining a span extending five plays into the future, the influence of that positive play on the performance score diminishes. And this works in the opposite direction for negative impact plays leading to the negative correlation. Notably, when the direct impact of the current play on ΔP is controlled for, the analysis reveals no significant correlation between the impact magnitude of a play and future performance. This finding suggests that the immediate impact of a play does not have a lasting influence on future performance.

Through analyzing two different approaches to quantifying momentum we found no evidence to suggest that momentum has an effect on tennis players' performance.

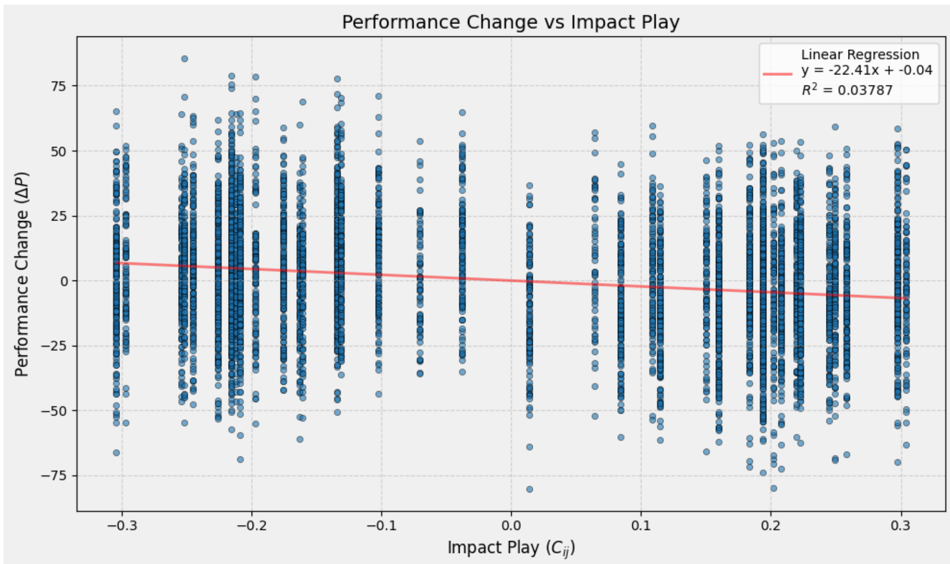


Figure 7: Shows the relationship between Performance Change and Impact Plays as defined in section 4.3 and 4.1, respectively

5 Predicting swings in the flow of play

5.1 Defining performance swings

Within a match, a state i is a *swing* in performance if the magnitude between its performance score and the performance score from state $i - 5$ is in the top 15% of all such differences. Note that swings are symmetric—a positive swing for one player is identical to a negative swing for the other, and vice versa—so we only classify the top 7.5% of magnitudes as swings in practice.

5.2 Predicting swings with Bayesian logistic regresion

Potential indicators

Unlike the score predictors from Section 2.1, which were defined with respect to the serving player in a single game, indicators for performance swings are defined for players across all games in a match. We fix "Player 1" and "Player 2" for an entire match for the following selected indicators:

- **Intercept (β_0).** The true log-odds of observing a swing in our given dataset.
- **P1 winning shot (X_1) and winning shot rolling sum (X_2).** Indicates when Player 1 hits an untouchable winning shot, and an ongoing count of untouchable winning shots hit in a single match.

- **P1 break point won (X_3) and break point rolling sum (X_4).** Player 1 wins a point that wins them the game, and an ongoing count of these break points.
- **P2 double fault (X_5) and double fault rolling sum (X_6).** Player 2 misses both serves and loses the point to Player 1, and an ongoing count of Player double faults in a match.
- **P2 break point missed (X_7) and rolling sum of missed break points (X_8.** Player 2 is the non-serving player and misses a point that would win them the game, and an ongoing count of winning points missed when Player 2 is not serving.
- **P2 unforced error (X_9) and unforced error rolling sum (X_{11}).** Indicates when Player 2 makes an error that was not forced by Player 1, and an ongoing count of unforced errors.
- **Fatigue difference (X_{11}).** Defined the same as in Section 2.1.

Model specification

We use the same general model for predictions as in Section 2, a **Markov chain** with **Bayesian logistic regression**. However, instead of specifying unique priors for each regression parameter, we use

$$\beta_0 \sim \text{Normal}(-2.513, 0.5)$$

$$\beta_1, \cdots, \beta_{11} \sim \text{Normal}(0, 20).$$

The prior for β_0 is determined by taking the true probability of observing a swing in our given dataset, then transforming that probability to log-odds. Choosing a weakly informative prior for all other coefficients allows the posterior distribution, and therefore the relative importance of each indicator, to be based on evidence from observed data.

Errors and uncertainties

Since our dataset is highly imbalanced—just 7.5% of points are swings—typical evaluation metrics are misleading. Indeed, our model predicts "no swing" on every point of real match data, but it still has 92.5% accuracy.

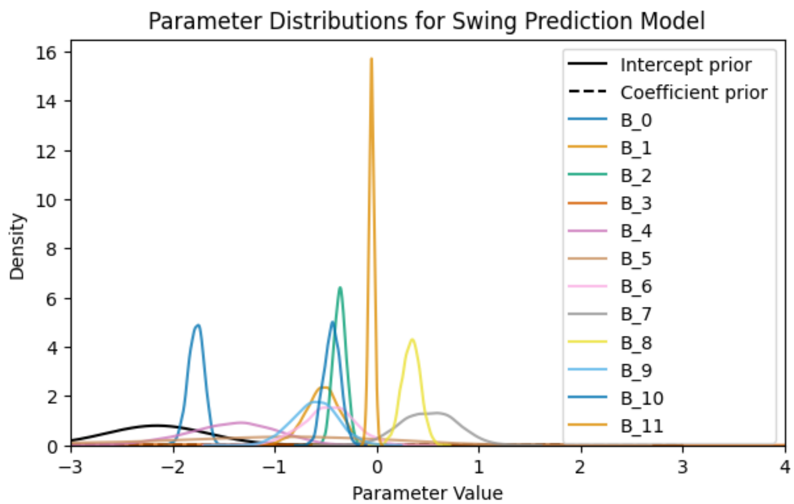
Sensitivity analysis

To test the strength of parameter priors, we double the mean μ_0 for the intercept β_0, and sample all other coefficients from Normal$(10, 20)$. Prior and posterior results of sensitivity analysis are shown in Figure 8. After adjustment, we see that the means have shifted to match the priors, and several distributions are bimodal.

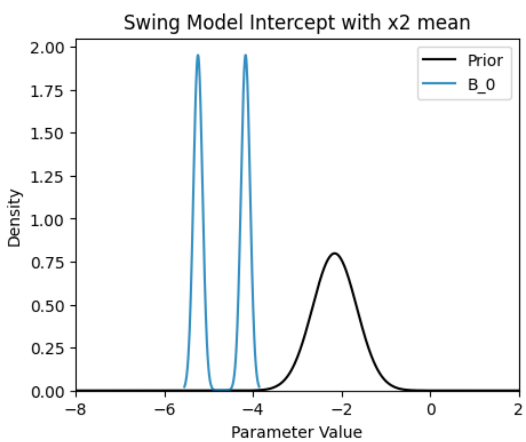
6 Generalizing the model

Binary outcomes are necessary

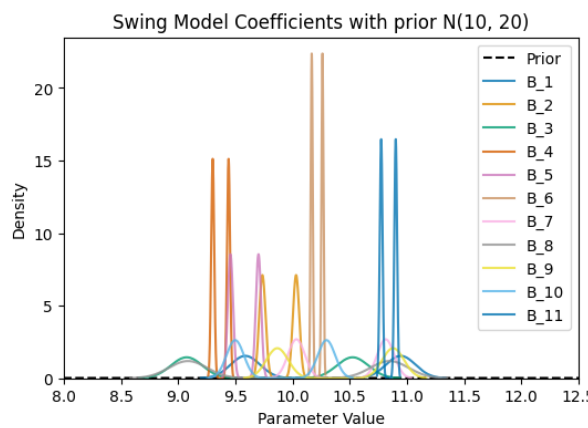
The Bayesian logistic regression model used in this paper is great for its adaptability in analyzing various attributes, making it a valuable tool for a wide variety of sports. This flexibility is very beneficial in



(a) Distribution of parameters using unmodified priors. We use a weakly informative prior of Normal$(0, 20)$ for all coefficients.



(b) Distribution of the intercept β_0 when sampled from Normal(-5.026, 0.5), a doubling of the original mean μ_0.



(c) Rublev vs. Goffin (third round, match 13, set 3, game 5). Rublev is the server and loses the match.

Figure 8: Prior and posterior distributions for the regression parameters used to predict swings in performance. Figures (b) and (c) show results of sensitivity analysis on the intercept and coefficients.

identifying key factors influencing performance, offering players and coaches insight into specific aspects of their sport. However, the model is contingent on the sport having binary outcomes, such as scoring a point or not. So, while refitting our model for table tennis and women's tennis would work, given we had the data, a sport like football or soccer would not. A primary challenge in using the logistic regression model lies in the collection and selection of relevant data. For instance, a table tennis coach interested in the impact of the ball's spin rate on serve outcomes could effectively use this model to find out how important spin rate is. It is suited for assessing whether a serve results in a score or not, or for predicting binary outcomes such as an ace versus a return. The key is the availability and quality of the data: with appropriate data, this model we presented can be very useful in quantifying the influence of specific attributes on the outcomes of a sport.

Performance score of the 2023 WTA Wimbledon final

The 2023 Women's Wimbledon finals saw an exciting match between Markéta Vondroušová and Ons Jabeur in which Vondroušová handily won both sets in the two out of three. In order to apply the performance score model to the game we needed point-by-point data which was retrieved from flashscore.com as well as an expected score for the server and the returner, respectively. We got the expected score from a study by (Carboch, Jan and Kočíb, Tomáš)[2] which included the rate at which women's doubles player's win the serve, because we could not find data for women's singles players. We assume this to be a sufficient proxy because the rate at which the men's doubles players scored the serve was only 2% difference from the rate at which the men's singles players scored the serve. With that rate we were able to calculate the performance score for the 2023 Women's Wimbledon Finals match, shown in Figure 9

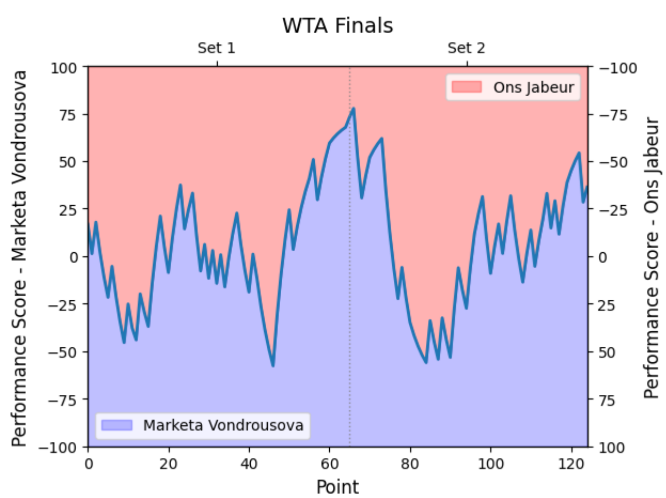


Figure 9: The performance score for both Markéta Vondroušová and Ons Jabeur in the 2023 Women's Wimbledon Finals match

6.1 Limitations and future directions

There are several limitations to our modeling approach that may be corrected in future adaptations:

- The logistic function may be too simple to fit to very imbalanced datasets, like our score outcomes (69% of scores are made by the serving player) and performance swings (92.5% of points do not represent a swing). On the other hand, a more powerful model that captures complex non-linear relationships is difficult to for players and coaches to interpret.

- Two of the score predictors, fatigue and speed, can only be known with precision after the game. However, players can still use our model results about the relationship between predictors and performance to guide in-game decisions intuitively.
- Some sensitivity analysis was conducted to test our assumption about the states being identical, but more fine-grained investigations could be done. Testing model performance on each state separately could be particularly informative.

7 Conclusion

We use a fundamentally Bayesian approach to model the flow of play within tennis games. Rather than making deterministic predictions about point outcomes, we generate probabilities—and probabilities of probabilities—using Bayesian logistic regression. Although our logistic model did not to extract useful evidence for momentum and prediction swing indicators from our skewed datasets, we believe this outcome itself can offer insights to coaches and players by guiding their training focus to more tangible skills, like building endurance and improving serving speed.

In the process of making predictions, we developed definitions of “performance” and “momentum” that mathematically implemented the most notable aspects of each: for performance, we incorporated dynamic updating using an exponential moving average; for momentum, we built upon our Markov model of game flow.

At each step of our modeling method, we considered the problem context—advising tennis coaches and athletes—when interpreting results. Our analysis naturally leads to many avenues for future research, from different ways to dissect a tennis match to alternative, mathematically rigorous definitions of momentum.

8 References

[1] *Atp elo ratings*. https://web.archive.org/web/20230622000010/https://tennisabstract.com/reports/atp_elo_ratings.html, 2023. Accessed: 2024-02-03.

[2] J. Carboch and T. Kočíb, *A comparison of service efficiency between players of male and female doubles at professional tennis tournaments*, AUC Kinanthropologica, 51 (2015), pp. 56–62, https://doi.org/10.14712/23366052.2015.32.

[3] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer Series in Statistics, Springer New York Inc., New York, NY, USA, 2001.

[4] E. Ma, *Markov Models From The Bottom Up, with Python - Essays on Data Science*, https://ericmjl.github.io/essays-on-data-science/machine-learning/markov-models/ (accessed 2024-02-04).

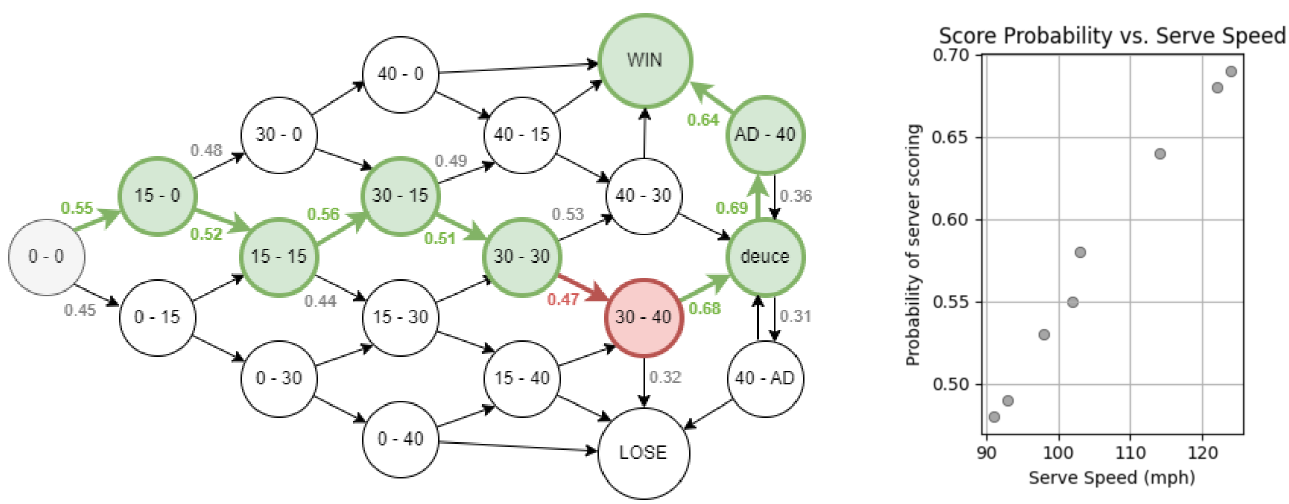
Memo to coaches

To coaches of professional tennis,

Have you ever wondered how your players should handle the pressures of “momentum” in a match? How to anticipate dramatic, exhilarating swings in performance, like those seen in the 2023 Wimbledon Championship final between Alcaraz and Djokovic? Or, perhaps, you’re skeptical that momentum has plays a role in game dynamics at all?

By analyzing data of men’s single matches from the 2023 Wimbledon Championships, our team has modeled the point-by-point flow of play, quantified relative performance and momentum, and shown how to predict swings of professional tennis matches. Using our mathematical definition of momentum, we find *no conclusive evidence* for an effect of momentum on player advantages and performance swings. Further, factors like serving speed, player Elo score, and the relative fatigue experienced by players do not entirely predict whether a player will score a point—the most important predictor remains whether the player is serving.

The finer insights of our analysis can be shown in our model for the flow of play, seen in Figure 10. Rather than predicting a single yes-or-no outcome for each possible score, we generate *probabilities* for each possible outcome. As the scatterplot in the figure shows, this method reveals that faster serve speeds improve a player’s chances of scoring. Although this fact feels intuitive, we believe that this model can be used to discover less obvious relationships between player attributes and scoring potential. This way, our model can be a valuable tool for guiding and supporting coaching strategies with empirical data.



(a) Diagram for the flow of a single game. Each node is a (`server score`, `receiver score`) state, and arrows are labeled with the probability of state *transitions*. Green nodes are correct predictions from our model and red nodes are incorrect predictions.

(b) In this game, faster serves are correlated with higher chances of the serving player winning the next point.

Figure 10: An application of the model to real game data. The probabilities between different scores highlight subtler factors that improve players’ chances of scoring, and can even be used to predict future scores.

As a part of our analysis, we defined a measurement for a player’s performance using an *exponential* *weighted moving average* that captures point-by-point variability, with 50% of the performance coming from the previous five points. This metric is scaled to be between -100 and 100 for ease of interpretation. For example, if the server is 72, then the receiver is -28, with -100 and 100 being the theoretical minimum and maximum, respectively.

We define momentum by the change in the probability of winning over the course of a match. To quantify momentum, we first determined the probability of winning a game from every state, with certain states having a higher impact than others. Then, we analyze the state-to-state transition shift of the probability because this shift indicates the importance of each point and quantifies the change in the likelihood of winning a game. The larger the change is, the more the point increases the probability of winning a game, and the smaller the change is, the less the probability of winning a game decreases. Then, we use an exponentially weight average for all previous points to highlight the significance of recent events over earlier ones. Nonetheless, to better understand the momentum and assess its effect on a player's performance, we look at the difference between six points. This approach allows us to pinpoint the locality of momentum as it occurs over the span of a match and capture the player's performance over an interval of six points, providing a comprehensive assessment of their play level.

When evaluating the relationship between the change in player performance and momentum, we found a negative correlation. However, this is not an indication that momentum adversely affects performance; rather, it suggests that if a player gains momentum by performing well, their performance eventually regresses to its mean value. We also found a negative correlation between the impact of a play—a transition between different states of the game—and momentum. Similarly to momentum, we concluded that a play's impact doesn't negatively affect performance but is an indicator that after a great shot, the player's performance regresses to its mean. This inference can be attributed to the discrete nature of the model used to calculate the performance score. In conclusion, we found no evidence indicating that momentum affects a player's performance throughout a match.

Based on our results, we recommend that coaches limit consideration of factors with contested definitions, such as momentum swings in a match. Instead, focus on tangible skills that make small improvements to your players' chances of scoring points, such as building endurance to combat fatigue and improving serve speed. Further, our model for the flow of play can be used to identify game situations that reliably alter probabilities of a score. The approach we suggest is subtle, as an outward change in scoring points is not immediately guaranteed. Nonetheless, we believe a steady approach of building up foundational skills will lead to improved playing consistency and an overall better experience than following unexpected swings of "momentum."

We hope our information and recommendations help you prepare your players to respond to events during matches and "tilt the flow of play in their favor."

Best regards,
Team 2429961