Journal of Electronics and Information Science (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/jeis.2025.100102
ISSN 2371-9524 Vol. 10 Num. 1

# *An Exploration of Ball Game Momentum Fluctuations Based on Multiple Regression Analysis and Convolutional Neural Networks*

**Yuqing Xia[1,*], Zhihan Gong[1], Fangyu Wei[1]**

[1]*School of Mathematics and Statistics, Northwestern Polytechnical University, Xi'an, China*
[*]*Corresponding author: gongzhihan@mail.nwpu.edu.cn*

***Keywords:*** ANOVA, Multiple Linear Regression Models, Hypothesis Testing, Convolutional Neural Networks

***Abstract:*** The aim of this study is to explore the patterns of player momentum fluctuations in tennis matches through a model combining multiple regression analysis and convolutional neural network (CNN). The study first analyzes real-time match data using analysis of variance (ANOVA) to identify the key factors that affect players' scores. Based on these factors, a multivariate regression analysis model is constructed for evaluating players' winning ability and their performance at specific moments. Then, the game data are deeply analyzed by convolutional neural networks to capture the fluctuating trends of player momentum. In addition, this paper verifies the non-randomness of momentum fluctuation by hypothesis testing method, which proves that the fluctuation of momentum in a match has a certain regularity and is closely related to the performance of players. The innovation of this study is that an analytical framework combining multiple regression and convolutional neural network is proposed, which not only improves the accuracy of momentum prediction, but also provides a new idea for dynamic analysis and optimal training of tennis.

## 1. Introduction

This paper synthesizes the algorithms of multiple linear regression model [1], analysis of variance [2], hypothesis testing [3] and convolutional neural network [4] to delve into the influencing factors of players' performances in tennis matches [5]. The study first analyzes players' scores in matches through multiple linear regression models to identify the key technical indicators that affect winning ability [6]. Then, the hypothesis testing method was used to verify the randomness of players' performance and explore the influence of “momentum” on the game results. In addition, a convolutional neural network model is constructed to predict the fluctuation of players' performance during the game to further improve the accuracy and applicability of the model. Through the comprehensive analysis of the data, this paper not only provides a scientific basis for optimizing players' techniques and strategies, but also lays a foundation for future game prediction and decision support.

## 2. Assessment of Players' Winning Ability and Performance

To address the initial question, this section performs a stepwise regression analysis using 13 technical statistical indicators of players. A multiple regression model was built using the players' scores in 127 matches as the dependent variable and the other 12 indicators as independent variables. At the same time, the study obtained unstandardized and standardized regression coefficients to assess the players' ability to win and to determine their performance at a given time.

### 2.1 Objective function

To build an appropriate scoring model based on the dynamic performance of the athlete in the model, a multiple linear regression model is introduced here. The multiple linear regression model can be expressed in the following form, i.e., the multiple linear regression model is introduced into the matrix representation.

$$\mathrm{Y}=\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \mathrm{X}=\begin{pmatrix} 1 & x_{11} & \cdots & x_{p1} \\ 1 & x_{12} & \cdots & x_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{pn} \end{pmatrix}, \mathrm{B}=\begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{pmatrix}, \varepsilon=\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \tag{1}$$

A stepwise regression analysis can be performed using the player's total score as the dependent variable and the remaining 12 indicators as independent variables to obtain the corresponding unstandardized regression coefficients and form the corresponding regression equations. When new data on game conditions are known, the strength can be evaluated according to the data.

The pseudo R-squared in the model is shown in the Table 1 below. It can indicate the degree of fit of the model.

Table 1: Pseudo R-squared table

| Pseudo R-square | |
|---|---|
| Cox Snell | .491 |
| Negolko | .491 |
| McFadden | .069 |

The model is built by making some assumptions about the ideal model. The chi-squared statistic is the difference in -2log likelihood between the final model and the simplified model. The simplified model is formed by omitting one effect from the final model. The original assumption is that all parameters of the effect are 0. Since omitting the effect does not increase the degrees of freedom, this simplified model is equivalent to the final model, which shows the feasibility of this simplified model.

Through programming analysis, it has identified several factors that have the greatest impact on the player's winning ability among the many factors that affect the player's score, and discarded other factors that have less impact. The complete model of a player's winning ability is shown below.

$$\mathrm{Y}=12.16+10.63X_1-14.797X_2+0.785X_3+0.846X_4 \tag{2}$$

Where $\varepsilon_i \sim \mathrm{N}(0,\sigma 2)$

### 2.2 Restrictive condition

Data analysis and calculation reveal the four key indicators that determine the winners and losers of the Men's Singles matches in the 2023 Wimbledon Open: second serve percentage, unforced errors, break point percentage, and reception point percentage.

By comparing the tennis variable winning factors we analyzed the effect of the percentage of serve

receive wins and losses on the match data of Carlos Alcaraz and Novak Djokovic in the Wimbledon Men's Open. It is important to note that this analysis is objective and free from bias. We analyzed the effect of the percentage of serve receive wins and losses on the match data of Carlos Alcaraz and Novak Djokovic in the Wimbledon Men's Open. We analyzed the effect of the percentage of serve receive wins and losses on the match data of Carlos Alcaraz and Novak Djokovic in the Wimbledon Men's Open. The percentage of winners and losers in serve reception of the two players was obtained through statistical analysis. Based on the statistics, the winning team's serve-receive percentage is 35.31%, while the losing team's is 28.14%, a difference of 9.17 percentage points. This suggests that the winning team has a higher scoring rate than the losing team, indicating that serve reception is a key factor in determining the outcome of the match.

### 2.3 Solution of the model

After processing and analyzing the known data, we can evaluate the strength of the tennis player to determine the performance of the player at a given moment. The course of the match is accurately described in real time.

In order to accurately measure a player's performance at a given moment, we visualize the player's on-court performance by plotting the function Y of a multiple linear regression model with respect to the duration of the match around the key winning factors of a tennis player, i.e. the percentage of second serves scored, the percentage of unforced errors, the percentage of breaks, and the percentage of points scored on the receiving serve. A player's on-court performance is affected by a variety of factors, such as whether he serves or does not serve, the number of service breaks, unforced errors on the court, and so on. We build a multivariate visual regression model to evaluate player performance more intuitively. Figure 1 shows the visualization.

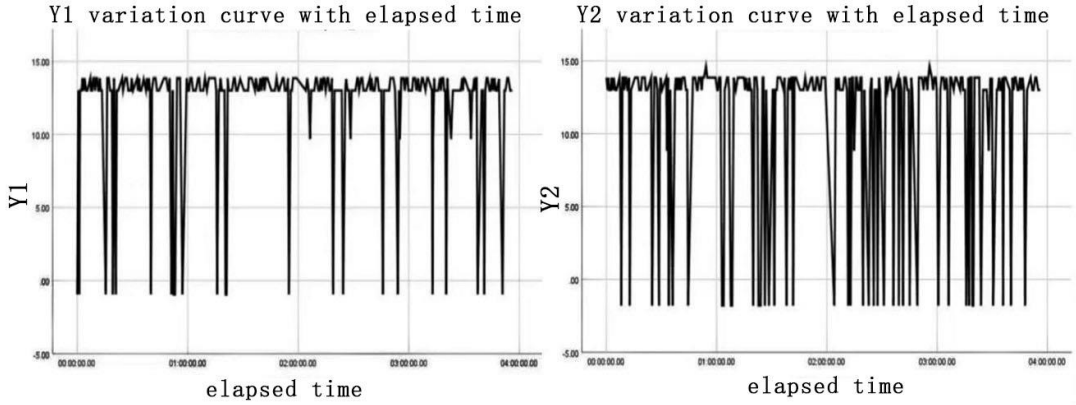


Figure 1: Multivariate visual regression visualization plots

## 3. Real-time Processing Analysis

### 3.1 Data processing

To select the appropriate hypothesis testing method, we need to process the data. Bayesian estimation is used to determine that the data follows a normal distribution, allowing us to choose the Z-test method for hypothesis testing. Part of the processing is shown in Figure 2.

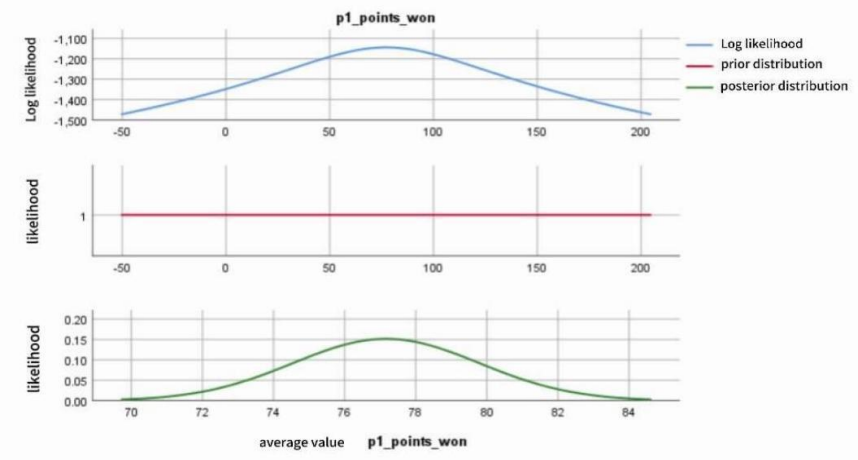


Figure 2: Analysis of results

## 3.2 Hypothesis testing

The Z-test was utilized to compare the difference between the sample and a known overall mean. This was done because men's singles tennis matches have rapidly changing conditions, involve a lot of data, the sample has a large capacity, and follows a normal distribution.

When using the Z-test, the first step is to calculate the value of Z using the formula.

$$Z = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \tag{3}$$

The next step is to compare the calculated Z-value with the theoretical Z-value to infer the probability of occurrence based on the relationship between the Z -value and the P -value. Then, the degree of difference can be assessed.

## 3.3 Test results

To accurately measure an athlete's momentum in a race, only a select few data points were analyzed, including speed mph, p1 distance run, and p2 distance run. Speed mph is analyzed specifically below.

For the speed mph variable, the Z-test variables include solving for the test sample mean, the known overall mean, the overall standard deviation, and the sample size of four values. This is shown in Table 2.

Table 2: Speed mph analysis

| Set no | average value | Number of cases | standard deviation |
|---|---|---|---|
| 1 | 114.97 | 62 | 11.402 |
| 2 | 116.95 | 92 | 11.778 |
| 3 | 115.76 | 54 | 12.011 |
| 4 | 116.35 | 81 | 11.427 |
| total | 116.13 | 289 | 11.607 |

We can use the Z-test to determine the value of Z based on the results presented in the Table 3 below. This will allow us to make a judgment on the degree of difference between the results of the player performance random and the results of the player affected by momentum. The degree of difference is significant.

Table 3: Relationship between Z-value and P-value

| \|Z\| | P-value | Degree of variation |
|---|---|---|
| ≥2.58 | ≤0.01 | remarkable |
| ≥1.96 | ≤0.05 | statistically significant |
| <1.96 | >0.05 | insignificant |

## 4. Predicting Fluctuations in Players' Winning Ability

### 4.1 Convolutional Neural Networks

This section utilizes a convolutional neural network for analysis, where the network is constructed by the exact number of training, testing, validation, and exclusion samples used, and the samples are carefully selected for training and testing.

Using convolutional neural network, it is possible to predict the fluctuation of the athlete's game. The neural network model has high feasibility due to the high accuracy of the model prediction.

### 4.2 Relevant factor

In this section, we analyze the winning factors of tennis players in matches, and after analyzing the stepwise regression analysis, eight indicators were excluded from the many factors. The remaining four indicators, i.e., second serve rate, unforced errors, break rate and serve receive rate, entered the equation successfully after the P-value test and were included in our consideration as the main factors affecting players' ability to win.

### 4.3 Model prediction

By describing and fitting trends to known data, the trends described are called "momentum", which allows further predictions of the performance of athletes on both sides of the game, and thus comparisons with future reality. Most of the time the fluctuations are within a relatively small range, with larger fluctuations at individual points. Most of the time the fluctuations are around the value 13, with smaller upward fluctuations and larger downward fluctuations.

### 4.4 Suggestion

Between the derived regression equations, the most critical winning factors in tennis can be analyzed based on the equations. It is possible for players to determine the importance of the winning factors based on these equations so that they can target their training and pay special attention to the winning points in these areas of the game, i.e., they cannot ignore the second serve percentage, the unforced error percentage, the percentage of breaks of serve in the match, and the percentage of serve-receive in the match.

## 5. Model Testing Analysis

To test the generalization of the model by understanding the data from the official website, data from the 2024 Australian Open men's singles tournament was chosen to be included in the model to further test the model. Initially, when constructing the model, the data was taken from the Wimbledon Open. If the model is found to have high predictive accuracy when tested with other match data, it can be hypothesized that the model will be more effective in predicting match fluctuations.

In the process of constructing the convolutional neural network, the neural network system was

trained in this section, and a total of 5350 valid samples were selected during the training process, which is many samples with high prediction accuracy.

The model has been tested to achieve 91.8% overall correct percentage of training and 92.3% overall correct percentage of testing for point victor, 86.1% overall correct percentage of training and 87.7% overall correct percentage of testing for point victor, and for point victor, the overall correct percentage of training is 98.6% and the overall correct percentage of testing is 98.9%.

After substituting Australian Open 2024 Men's Singles match data, we tested the model, which resulted in the following images about the predicted probability of fitting, part of the display is shown in Figure 3 below.

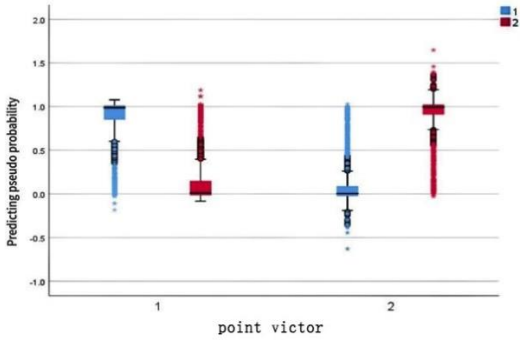


Figure 3: The predicted proposed probability for point victor

## 6. Conclusion

This paper combines convolutional neural network (CNN) and multiple regression analysis (MRA) modeling to investigate the patterns of player momentum fluctuations in tennis matches and their effects on match results. Through analysis of variance (ANOVA), this paper identifies the key factors affecting players' scores, such as second serve rate, unforced errors, break rate, and serve-receive rate, and establishes a multiple regression model to assess players' winning ability. Hypothesis testing analysis showed that the fluctuation of players' performance is not random, but influenced by "momentum". The model was further validated through convolutional neural network analysis, and the residual analysis showed that the predictions were in good agreement with the actual matches, indicating that the model has a high accuracy. The study also found that technical indicators such as percentage of second serves, unforced errors, and percentage of breaks of serve had a particularly significant effect on match fluctuations.

## References

*[1] Liu Changyu. Optimal experimental design for the comparison of regression model curves [D]. Shanghai Normal University, 2024.DOI: 10.27312/d.cnki.gshsu.2024.000172.*

*[2] Fan Jiangyuan. Analysis of variance and covariance matrix function of multivariate functional type data testing problems and applications [D]. Huaqiao University, 2023. DOI: 10. 27155/ d.cnki.ghqiu.2023.000636.*

*[3] Xu Jianjun. Research on hypothesis testing and sufficient downscaling in functional regression models[D]. University of Science and Technology of China, 2024. DOI: 10. 27517/ d.cnki.gzkju.2024.000004.*

*[4] Qi Zhankui, Zhang Xinpeng, Liu Xuliang, et al. A one-dimensional convolutional neural network-based intelligent identification method for well test model [J]. Oil and Gas Well Testing, 2024, 33(02): 72-78.DOI: 10.19680/j.cnki.1004-4388.2024.02.012.*

*[5] Zhang Peixiao. Construction of tennis match result prediction model[J]. Science and Technology Innovation, 2024, (19):33-36.*

*[6] Balasubramanian A G, Vinuesa R, Tammisola O. Prediction of flow and elastic stresses in a viscoelastic turbulent channel flow using convolutional neural networks[J]. Geophysical Journal, 2024, V41(4).*