An analysis of the hot hand phenomenon in basketball and mid-range shooting

Haoran He

Department of Statistics, University of Pittsburgh, Pittsburgh, 15213, The United States

q56586668@gmail.com

Abstract. Does the Hot Hand phenomenon exist during basketball, especially in the NBA? This question has been controversial over the past years in the sports field. The enormous literature on the Hot Hand effect in basketball was conducted to investigate the Hot Hand hypothesis. This study analyzes this question using a novel data set of all mid-range shots from the 2023-2024 NBA regular season, combined with data on teams and players. A model was built based on the definition of the Hot Hand and potential influential variables such as distance and location to analyze the players' shot data. Supervised machine learning methods fit the model and measure its performance. The results suggest that mid-range shooting streaks do not affect making the next shot in game situations, indicating that the Hot Hand effect is not present in mid-range shooting situations.

Keywords: Hot Hand, Basketball, 2023-2024 NBA regular season, Mid-Range Shot.

1. Introduction

The "hot hand" (also known as the "hot hand phenomenon" or "hot hand fallacy") is a previously considered cognitive social bias that suggests a person who has achieved success in a particular task is more likely to be successful in future attempts. This concept is commonly associated with sports and skill-based activities and originally comes from basketball, where a player is believed to have a higher chance of scoring if they have made successful shots in succession, known as having the "hot hand".

The key point at the center of research into the hot hand is whether the widespread belief that previous shooting streaks increase a player’s chance of hitting the next shot is true. If a NBA player makes two, three or four shots in a row, then he is hot handed and is more likely to make his next shot than expected. This study started with Gilovich et al. [1], who demonstrated that the phenomenon was a cognitive illusion caused by random sequences. For example, when a person flipped a coin ten times, there was a chance of having five consecutive heads. However, it was difficult to conclude that the previous four heads increased the chance of getting a head in the fifth attempt. Thus, Gilovich et al. provided that there was no sufficient evidence for a correlation between successive shots and the chance of making the next shot. Then, William O. Brown and Raymond D. Sauer [2] set a point-spread pricing model to prove that belief in the hot hand affected the point-spread betting market, but their study didn't support the hot hand in a real-world context.

Recent research into the hot hand has usually concentrated on controlled settings such as shooting experiments, the NBA 3-point contest, 3-point shooting, and free-throw shooting in games. Both Arkes

[3], Yaari and Eisenmann [4] found evidence of the hot hand phenomenon in free throws. Miller and Sanjurjo [5-7] demonstrated the hot hand in controlled and semi-controlled settings in their three papers. At the same time, recent experiments have tested for a hot hand in the run of play in NBA games. Bocskocsky et al [8], and Csapo and Raab [9] found that both offences and defences react to made shots. In the run of play, Lantis and Nesson analyzed detailed data on free throws [10] and the NBA 3-point contests [11]. They found a small hot hand effect for free throws and 3-point shots within shot locations, while they had the opposite results for field goal attempts across shot locations. Studies have continued to follow up and revived the hot-hand debate in academia. Kostas Pelechrinis, an associate professor of computing and information at Pitt's School of Computing and Information, claimed that Hot hand exists [12] and "players can indeed get hot in actual live-game situations [13]."

The study aimed to empirically analyze whether the hot hand exists in the NBA for mid-range shots. The author used detailed data for the 2023-2024 NBA regular season. The data from modern-era basketball provides a large sample size and better quality in terms of specificity relative to the datasets used before. Thus, this project expects a better outcome and interpretation of hot hand with a modern dataset of shooting records of renowned NBA players. Moreover, previous papers only considered 3-point shots and free throws in games, but they ignored mid-range shots, which were 2-point shots and more likely to reflect the hot hand effect. Therefore, this paper focused on an analysis of the hot hand effect for the mid-range shots.

2. Data and Methods

2.1. Data

2.1.1. Data Overview

The main data source consists of play-by-play data for the 2023-2024 NBA regular season from the nbastatR package, which is downloaded from Github. This package combines data from credible sources, including NBA's API, HoopsHype, nbadraft.net, and Basketball Reference.com. One of the datasets provided by the package is game_logs, which contains a history of in-game events. For the 2023-2024 regular season, the game_logs data consists of 58 columns and 26,401 rows. This dataset is used to identify the NBA player with the highest-scoring game, as these players are more likely to have hot hands. The top 5 players who scored the most in a single match were then identified. Additionally, the shot data for five players includes 1487, 1436, 1305, 1652, and 851 rows, respectively, with each player having 27 columns of shot-relevant features such as location coordinates, shot types, and whether the player made the shot or not.

This paper also used another data set called team-shots, which provided the teams and players for every game and detailed information for every event in each game, including typeEvent, typeAction, typeShot, zoneBasic, isShotMade, and so on.

When the author set zoneBasic = "Mid-Range", the team-shots data set could build the mid-range shooting data set of each team and each player. Thus, after data wrangling, the mid-range shooting data set was used to test the existence of the hot hand fallacy.

2.1.2. Exploratory Data Analysis

The study utilized histograms to visually represent the distribution of game scores and players. The goal was to identify players with high scores in a single game, as they were more likely to have "hot hands". The histogram showed that the scores from NBA players followed a right-skewed distribution, with only the NBA All-Star roster scoring more than 25 points in the 2023-2024 regular season. Players such as Luka Doncic, Joel Embiid, Shai Gilgeous-Alexander, Kevin Durant, and Devin Booker were identified as the most likely "hot-hand" mid-range shooting players.

Subsequently, shot charts were used to analyze the shot location and shot types for these players. The analysis revealed that Luka Doncic was likely to make his mid-range jump shots from the left side and left center areas, while Kevin Durant attempted a significant volume of successful mid-range shots in

DOI: 10.54254/2755-2721/87/20241597

the right side and right center areas. Joel Embiid tended to make a majority of jump shots from the high post area, while Shai Gilgeous-Alexander favored mid-range attempts from both wings and near free throw line. Devin Booker was a phenomenal mid-range shooter, exhibiting a high shooting percentage across all areas.

The study utilized Coxcomb charts to illustrate the predominant shot types for each player and their respective proportions. The visual representation indicated that Luka Doncic showed a preference for step-back jump shots and a few turnaround fadeaway shots, while Joel Embiid favored jump shots and pull-up jump shots. Shai Gilgeous-Alexander tended to favor pull-up jump shots and step-back jump shots, while Kevin Durant and Devin Booker displayed a notable tendency for pull-up jump shots.

2.2. Methods

2.2.1. Variables and Model Specification

To test the hot hand, this study built a logistic regression probability model that could represent the definition of Hot Hand. Incorporating these factors as additional controls in the regression was crucial for understanding the impact of other variables on shot success. Without these factors, there could be a risk of omitted variable bias. Furthermore, regression analysis allowed for a natural investigation of subsets of shots to uncover whether a "hot hand" effect was concentrated on specific shot locations or distances. Lastly, through regression, this study was able to adopt a flexible approach for measuring multiple streaks of success over the previous mid-range shots.

This model was shown in the following specification:

$$\Pr(\text{isShotMade} = 1) = \text{logit}^{-1}(\alpha + \beta 1\text{lastShot} + \beta 2\text{lastFive} + \gamma 1\text{locationX} + \gamma 2\text{locationY} + \gamma 3\text{distanceShot}) \quad (1)$$

where $\text{logit}(p) = \ln(p/1 - p)$, p was the probability that the player made the current shot, α was the intercept, β was slope for variables of interest, γ was the coefficient for the control variables.

Using this model, the author examined the effects of the previous mid-range shot on the probability of making the current mid-range shot.

In relation to the Hot Hand phenomenon, the response variable "isShotMade" was created to denote whether the player successfully made the attempted shot, which was a categorical variable The predictor variables "lastShot" and "lastFive" were established to capture the player's performance, with "lastShot" indicating the player's success in their previous shot and "lastFive" representing the average shot percentage of the last five shots. Other controlled variables considered in this study encompassed location, distance, name zone, and type of action. Ultimately, the selected covariate variables were "locationX" and "locationY," which signified the precise shot location, and "distanceShot," which denoted the distance of a shot made.

2.2.2. Methodological Approach

The statistical analysis was conducted using R version 4.4.0 and RStudio version 2024.04.0-735. The analysis focused on NBA single-game leaders and records for points during the 2023-2024 regular season, as these NBA players were more likely to make consecutive shots. Using R code, the top single-game leaders for points during the 2023-24 season was identified. Subsequently, exploratory data analysis, such as creating a histogram, was used to describe the game scores from the players.

Before testing the hot hand hypothesis, this paper analyzed each player’s shot preference, which indicated their preferred shooting locations on the court. Numerous studies have explored the relationship between shot locations and the hot hand phenomenon. According to Lantis et al. [11], "the difference in shot location from the previous shot also grows in magnitude for longer streaks of success." The findings from studying the hot hand and shot preferences could lead to further insights. For example, if a player was found to have a hot hand and is likely to shoot near the post, this could prompt further investigation into other players who tend to shoot from similar locations.

To understand how consecutive shooting streaks affect the next shot, this paper analyzed the likelihood of making the next shot based on the previous shooting streak. The fluctuation of the line plot didn't prove the existence of the hot hand. However, if the hot hand does exist, these players would experience an increase in their field goal percentage as they continued to make shots successively.

To test the hypothesis of the Hot Hand, the author built a logistic regression model to examine the association between the current shot and the previous shots. Using the summary() function, the author obtained regression coefficients which encompass estimated coefficients, standard error, z-value, and p-value.

This paper tried to predict the categorical response variable, isShotMadeBinary, using supervised learning algorithms.

First, the author divided the dataset into a training set comprising 75% of the original data and a testing set comprising 25% of the original data. The author then applied some classification methods including regularized logistic regression, random forest, and K-nearest neighbor (K-NN) to train the model. After that, this paper performed model tuning to identify the best parameters for each model. The author used 10-fold cross-validation to determine the appropriate hyper-parameters for each model. For regularized logistic regression, the author optimized the lambda parameter, while the author used grid search to find the best number of randomly selected predictors for the random forest algorithm. The author set the mtry hyper-parameter range from 2 to 16 and found the optimal value for each model. The author set ntree to 1,000 as a large number is generally better. Finally, the author compared the performance of each model by calculating the misclassification rate on the 25% testing set, and predicted the outcome for the testing set.

3. Results

This paper began by examining a simple hypothesis of the hot hand: whether a player who made his last shot was more likely than expected to make his next shot. Calculating the probability of making the next shot based on the previous shooting streak. This paper used the mutate function in R to create new variables: nowShot, previousShot, and shotPercentage. The variable nowShot meant the current status of consecutive shots, the variable previousShot meant the previous status of consecutive shots, and the variable shot percentage was the shooting percentage of shots made. Then the author created line charts for different NBA players to represent shooting percentage change over time. The fluctuation of line charts told us that players who made their previous mid-range shot were between 4 and 10 percentage points more likely to make their next mid-range shot. After making three or four consecutive shots, the shooting percentage would be at a stable level, which was between 8 and 10 percentage higher than expected. The results from these line plots did not provide sufficient evidence to conclude that the hot hand effect existed during the game of basketball. However, it gave reasonable and valid evidence to dive into a deeper analysis of the hot hand effect using the 2023-2024 regular season data set.

Next, this paper explored the logistic regression model for the hot hand effect. The results were shown in Table 1.

Table 1. Logistic regression analysis for the Hot Hand effect

Players	Variables	Summary			
		Estimated Coef	Std. Error	z value	Pr(>|z|)
Luka Doncic	(Intercept)	0.898855	2.572995	0.349	0.727
	lastShot	-0.095912	0.877982	-0.109	0.913
	lastFive	-1.045906	1.939354	-0.539	0.590
	locationX	0.001329	0.003828	0.347	0.728
	locationY	0.026732	0.017937	1.490	0.136
	distanceShot	-0.258224	0.271811	-0.950	0.342
Joel Embiid	(Intercept)	1.992510	2.157415	0.924	0.3557

Table 1. (continued).

	lastShot	0.563959	0.551546	1.023	0.3065
	lastFive	-0.870642	1.264764	-0.688	0.4912
	locationX	-0.010691	0.004732	-2.259	0.0239
	locationY	0.021724	0.012098	1.796	0.0725
	distanceShot	-0.317614	0.202069	-1.572	0.1160
Shai Gilgeous-Alexander	(Intercept)	0.674901	1.946527	0.347	0.729
	lastShot	-0.465617	0.626987	-0.743	0.458
	lastFive	-0.314106	1.694400	-0.185	0.853
	locationX	-0.001947	0.002704	-0.720	0.472
	locationY	-0.003746	0.006890	-0.544	0.587
	distanceShot	0.020757	0.147214	0.141	0.888
Devin Booker	(Intercept)	5.926E-01	1.127E+00	0.526	0.5990
	lastShot	8.236E-01	4.437E-01	1.856	0.0634
	lastFive	-1.113E+00	8.987E-01	-1.238	0.2157
	locationX	-8.534E-04	1.891E-03	-0.451	0.6518
	locationY	9.812E-06	4.546E-03	0.002	0.9983
	distanceShot	-3.711E-02	8.641E-02	-0.429	0.6676
Kevin Durant	(Intercept)	-1.346623	1.047587	-1.285	0.199
	lastShot	0.227615	0.372767	0.611	0.541
	lastFive	-0.218845	0.801211	-0.273	0.785
	locationX	0.001261	0.001578	0.799	0.424
	locationY	-0.006118	0.004118	-1.486	0.137
	distanceShot	0.125982	0.080553	1.564	0.118

In Table 1, the paper shows that Luka Doncic and Shai Gilgeous-Alexander had negative coefficients for lastShot and lastFive. Conversely, Joel Embiid, Kevin Durant, and Devin Booker had a positive coefficient for lastShot and a negative coefficient for lastFive. This means that Luka Doncic and Shai Gilgeous-Alexander were not influenced by the hot hand effect, while Joel Embiid, Kevin Durant, and Devin Booker were more likely to make their next shot after their last shot. However, it was also found that the hot hand effect was not significant since all P-values were greater than 0.05.

The paper used supervised learning methods to train the model. The author measured the model's performance by calculating the misclassification rate of logistic regression on a 25% testing set. The misclassification rates ranged between 27% and 57%, showing that the model's performance varied across different testing sets. This result suggested that the hot hand effect for mid-range shooting might exist for some NBA players in specific games, but it wasn't a consistent or significant factor in the NBA.

4. Conclusion

After analyzing the 2023-2024 NBA regular season data, this paper found that the "Hot Hand" phenomenon could not be a consistent or significant factor in the NBA. The statistical analysis all showed that factors like distance and location didn’t play a significant role in predicting and interpreting players’ next shot.

However, the concept of the "Hot Hand" is still complex. While some statisticians argue that it's a myth, there are other factors at play, such as muscle memory, psychological elements and defensive intensity, that haven't been fully accounted for in this paper. Further research should be conducted to

DOI: 10.54254/2755-2721/87/20241597

take these additional factors into account and develop more comprehensive methodologies to conceptualize player performance when taking a shot.

References

[1] Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. Cognitive psychology, 17(3), 295-314.
[2] Brown, W. O., & Sauer, R. D. (1993). Does the basketball market believe in the hot hand? Comment. The American Economic Review, 83(5), 1377-1386.
[3] Arkes, J. (2010). Revisiting the hot hand theory with free throw data in a multivariate framework. Journal of Quantitative Analysis in Sports, 6(1).
[4] Yaari, G., & Eisenmann, S. (2011). The hot (invisible?) hand: can time sequence patterns of success/failure in sports be modeled as repeated random independent trials?. PloS one, 6(10), e24532.
[5] Miller, J. B., & Sanjurjo, A. (2014). A cold shower for the hot hand fallacy (No. 518). IGIER working paper.
[6] Miller, J. B. and A. Sanjurjo (2018). Surprised by the gambler's and hot hand fallacies? A truth in the law of small numbers. Econometrica 86(6), 2019–2047.
[7] Miller, J. B., & Sanjurjo, A. (2021). Is it a fallacy to believe in the hot hand in the NBA three-point contest?. European Economic Review, 138, 103771.
[8] Bocskocsky, A., Ezekowitz, J., & Stein, C. (2014). Heat check: New evidence on the hot hand in basketball. Available at SSRN 2481494.
[9] Csapo, P. and M. Raab (2014). "Hand down, man down." analysis of defensive adjustments in response to the hot hand in basketball using novel defense metrics. PloS One 9(12), e114184.
[10] Lantis, R., & Nesson, E. (2021). Hot shots: An analysis of the "hot hand" in NBA field goal and free throw shooting. Journal of Sports Economics, 22(6), 639-677.
[11] Lantis, R., & Nesson, E. (2024). The hot hand in the NBA 3-point contest: The importance of location, location, location. Journal of Sports Economics, 15270025231222630.
[12] Pelechrinis, K., & Winston, W. (2022). The hot hand in the wild. PloS one, 17(1), e0261890.
[13] Pelechrinis, K (2024). How real is the 'hot hand' in basketball? https://fox8.com/news/nexstar-media-wire/the-hot-hand-is-a-real-basketball-phenomenon-but-only-some-players-have-the-ability-to-go-on-these-basket-making-streaks/