The 'Hot Hand' An Investigation into Streakiness in Shooting

The 'Hot Hand'

An Investigation into Streakiness in Shooting

a thesis presented
by
Max Bobby
to
The Departments of Statistics and Computer Science

in partial fulfillment of the requirements
for the degree of
Bachelor of Arts
in the subject of
Computer Science and Statistics

Harvard University
Cambridge, Massachusetts

April 2022

Thesis advisor: Alex Young Max Bobby

The 'Hot Hand'

Abstract

The concept of the 'hot hand' is highly debated in the fields of statistics, data science, and psychology. The overwhelming consensus for the past four decades has generally revolved around the concept that 'hot handedness' is nothing more than a misinterpretation of randomness in small sample sizes. More recent research indicates that these conclusions, specific to the shooting performance in basketball, were founded on flawed analyses, and that a 'hot hand' may indeed exist. However, researchers have yet to successfully identified the 'hot hand' in modern NBA game data, and primarily focus on shooting data obtained from controlled experiments.

In this work, we investigate streakiness in the context of player shooting data. Specifically, we develop a novel approach to the definition of streakiness that accounts for player specific effects. We then use two multivariate frameworks in the form of logistic regression and random forests, in order to measure the significance of streakiness in the context of predicting the outcome of a field goal attempt. Unlike previous research, we also examine this definition of streakiness with both pooled models and individual player-specific models. The results, similar to extant literature conflicted. Although the random forest models do not identify streakiness as a significant predictor, the logistic regressions provide several interesting insights. In particular, we find that while being 'hot' is significant at the .05 level in the pooled model, it is only inconsistently significant in the player-specific models. This indicates that individuals are likely to experience a 'hot hand' effect differently, if such an effect exists.

Contents

Listing of figures

List of Tables

This thesis is dedicated to my wonderful family and friends, who have never failed to stop supporting me in my endeavors.

Acknowledgments

I want to express my deepest gratitude to Alex Young for his wonderful advice and guidance through this process. I am also immensely grateful to my computer science and statistics professors throughout my time at Harvard for the enriching education that inspired me to explore this topic.

Thank you to my friends for supporting me through countless long nights working on problem sets, study sessions, and this thesis. You are the reason I have loved my time here, and you all continually inspire me to be the best version of myself.

To my parents and my sister, thank you for over two decades of endless support and love. I am eternally grateful for you all.

I never looked at the consequences of missing a big shot ... When you think about the consequences you will always think of the negative result.

Michael Jordan

1 Introduction

On January 23^{rd}, 2015, Klay Thompson scored 37 points in a single -quarter. He did so by shooting 13-13 from the field and 9-9 from beyond the arc. This inhuman performance was practically unheard of and nearly impossible from the perspective of many of the greatest NBA players [19]. Time and time again, the league's most brilliant players have 'shot the lights out' to the point that teammates and coaches alike recognize when a certain player is getting 'hot.' More recently, Steph Curry (who is perhaps uncoincidentally, Thompson's teammate and fellow 'Splash Brother'), made 105 three-pointers in a row at a practice. To put the unlikeliness of this event's occurrence into perspective, even if we were to suppose that Steph had an unrealistically high three-point shooting probability of 80%, if he were to take 500 consecutive shots in a row, the probability of experiencing a streak of similar length is less than 1 in 200 million. Moreover, even if he took 500 three points shots in a row, every day for the past 20 years, the probability of achieving a streak of comparable length is 0.000039 [5].

The concept of the ‘hot hand’ has been a hotly debated topic in the fields of statistics and psychology for several decades. Perhaps the most well-regarded study in this area was conducted by Gilovich, Vallone, and Tversky (GVT) in 1985. In their paper, “The Hot Hand in Basketball: On the Misperception of Random Sequences,” GVT sought to investigate the existence of a hot hand using data from the Philadelphia 76’ers shooting records (who at the time, were the only team to maintain accurate shooting data!), free throw records from the Boston Celtics, and a controlled shooting experiment with Cornell varsity basketball teams. Ultimately, GVT found no evidence of a ‘hot hand’ in any of the aforementioned data sets, leading them to classify any detection of streaks as a ‘general misconception of chance’ (Gilovich, Vallone, and Tversky, 1985) [8].

Notably, many players and coaches alike spoke out against the ‘hot hand fallacy;’ when Red Auerbach, a 16-time NBA champion coach, read the results, he reportedly said, ”Who is this guy? So he makes a study. I couldn’t care less.” Unfortunately for Auerbach (and any like-minded thinkers), for some time, there was little statistical evidence to support an argument for a hot hand — most research only further supported the hypothesis posited by GVT [11].

The fascination with the ‘hot hand’ is not purely limited to basketball. Researchers have also investigated the existence of a hot hand effect in a variety of other sports. Arthur et al. analyzed the existence of a hot hand effect using modern MLB pitching data on fastball velocities. Ultimately, Arthur et al. were able to leverage a Hidden Markov Model to predict when a player might be hot or cold. Moreover, they were able to determine whether a particular pitcher would be able to maintain a hot streak, or whether they would go cold. Indeed, using only two months of MLB pitching data, their model was able to better predict the velocity of future pitches better than the pitcher's average velocity across the remaining 6-month season. Perhaps most interestingly, they were able to find that if a pitcher was 'hot,' batters were much less likely to make hits or gain extra bases [2].

Otting et al. also explored the possibility of a 'hot shoe' effect in soccer within the context of penalty shots based on data from the German Bundesliga (one of Europe's premier soccer leagues). They utilized regularized Hidden Markov Model to model the latent forms of the players, incorporating the heterogeneous skill levels of penalty takers and goalkeepers along with a LASSO penalty. Ultimately, their results indicated that different forms of players may be tied to different states in the HMM, which provides evidence for the existence of a 'hot shoe' effect [20].

1.1 Review of Existing Literature

In recent years, researchers have discovered new evidence for the hot hand in basketball. Of particular interest is a paper written by Miller and Sanjurjo (2018) that identifies a flaw in GVT's statistical analyses. Specifically, they proved that for finite sequences of i.i.d. binary data, the proportion of successes following a streak of successes occurs with a probability less than the overall probability of success for the sequence. A simple example of this bias can be seen in Table 1.1.1.

After correcting for this "streak selection bias" in GVT's original data sets, Miller and Sanjurjo were able to find statistically significant evidence for a 'hot hand effect.' Of particular interest was the fact that in addition to detecting a hot hand effect at the individual level,

Sequence	Proportion of H's that follow one or more heads
TTT	-
TTH	-
THT	0
HTT	0
THH	1
HTH	0
HHT	$\frac{1}{2}$
HHH	1
Expectation :	$\frac{5}{12}$

Table 1.1.1: An illustration of the bias identified by Miller & Sanjurjo. [12]

Miller and Sanjurjo also found a pooled hot-hand effect across the entire sample. Indeed, they found an average hot hand effect across all players that represented a substantial increase in terms of field goal percentages. That being said, the effect was not heterogeneous across players, indicating that there is an incentive for coaches and players alike to identify which players may experience a hot hand effect [12].

Miller and Sanjurjo continued to explore the potential of a hot-hand effect in a variety of different contexts. In 2015, they investigated the existence of the hot-hand effect in the NBA 3-point context. Generally speaking, the NBA 3-point contest is possibly one of the most ideal settings for studying the hot hand effect, as the environment mitigates any confounding factors that might arise in a typical NBA game. In the 3-point contest, players are given 60 seconds to shoot 25 balls from 5 different positions along the 3-point line (in other words, they shoot 5 balls from each position). This effectively eliminates variability in distance, defensive pressure, offensive opportunities, etc., while still preserving a comparable amount of pressure to a typical NBA game. In this paper, Miller and Sanjurjo leveraged nearly 34 years of three-point contest data and applied a correction for the streak selection bias they identified in their first paper. Ultimately, Miller and Sanjurjo were able to select a number of players that experienced a hot-hand effect. Interestingly, similar to their first experiment, they also found a pooled hot-hand effect, despite heterogeneity in the magnitude and sign of the effect across players. Indeed, on average, Miller and Sanjurjo found that players shot anywhere from +5 to +9 percentage points better on a shot following a streak of hits [13].

Miller and Sanjurjo also conducted an improved field experiment similar to that of GVT conducted in 1985. This study had two phases, one in which players made only shots, and a second (6 months later) in which players placed bets on their shots before making them — in order to identify whether players could predict their own 'hot-hand' effect. Miller and Sanjurjo's study design improved upon GVT's in numerous ways. Phase one of the study allowed players to focus on purely shooting, whereas GVT potentially interrupted the rhythm of shot attempts by asking players to place bets between each shot. More importantly, players were asked to shoot from the same location, rather than being forced to move after each shot attempt. Finally, Miller and Sanjurjo also collected substantially more data, in that participants took 300 shots, as compared to 100 in GVT. Overall, the results yielded a strong hot hand effect ($>$ 9 percentage points) in certain individuals. Similarly to previous papers, they were also able to find a pooled hot hand effect across all players. Moreover, Miller and Sanjurjo found further evidence for the hypothesis that individuals and teammates may be able to identify and exploit tendencies to become 'hot' [15].

Other researchers have also investigated the 'hot hand' effect in controlled settings. Lantis and Nelson investigated the hot hand using NBA data on both free throws and field goal attempts. Overall, they found a small but persistent hot hand effect in free throws, with a magnitude of two to four percentage points. They also found that there was no hot hand effect associated with field goals, and that if a player makes three field goal attempts in a row, they are actually less likely to make their next field goal attempt by 0.6 percentage points. Notably, the paper does not account for the bias identified by Miller and Sanjurjo, indicating that the small hot hand effect identified in free throw shooting, as well as the insignificant hot hand effect in field goals could be more substantial in magnitude [9].

Arkes (2010) also examined the hot hand effect in the context of free throw shooting. Most prior studies tested for the effect with a univariate framework, which may have lacked the sufficient statistical power to detect the 'hot hand.' As a result, Arkes attempted to utilize a pooled logistic regression framework with individual fixed-effects in order to increase statistical power, while also controlling for separate players' abilities. Ultimately, Arkes found that if a player hits the first free throw, they are 2-3 percentage points more likely to hit the second free throw [1].

Further research has been completed on perceptions of the 'hot hand.' In their 2017 paper, Miller and Sanjurjo found that experienced players were actually capable of predicting their own 'hot hand' effect, a corollary of their original 2014 article focusing on the section of GVT's paper in which professional basketball players bet on themselves. After correcting for the streak selection bias, they found expert players were capable of predicting shot outcomes at a rate better than expected. They theorize that this may have come as a result of players' belief in the hot hand — in other words, because of their belief in the hot hand, players tend to automatically place bets after successes [14].

Interestingly, other studies have found a similar result in that players are more willing to 'bet' on themselves when they are on a streak, and take more risky shots following a series of made field goals. Csapo et. al. elected to measure NBA shot difficulty by three different metrics, shot distance, shot type, and shot angle. They found that the outcome of previous field goals had a significant effect on shot selection. If a player made more consecutive makes, shot difficulty increased with regards to the three aforementioned metrics. Similarly, with consecutive misses, players took less difficult shots. The results also implied that performance may actually improve during hot streaks, as shooting accuracy did not appear to decline, while shot difficulty increased [6].

Not all research agrees with Miller and Sanjurjo's findings. Daks, Desai, and Goldberg employed Miller and Sanjurjo's methodology to further examine the hot hand effect, but this time in the context of a few specific players, Stephen Curry, Klay Thompson, and Kevin Durant. Daks et al. also utilized modern NBA data from the 2016-17 regular season. In order to correct for the streak selection bias, Daks et al. performed 10,000 permutation tests on each game-long set of data for each player, and computed the following test statistic:

$$t_k = t_{k,hit}(X) - t_{k,miss}(X)$$

wherein X is a binary string of game data wherein '0' represents a missed field goal, and '1' represents a made field goal, k is the length of the streak, $t_{k,hit}$ which is the conditional fraction of hits given k prior hits, and $t_{k,miss}$ is the conditional fraction of misses given k prior misses. This can be thought of as the proportion of hit streaks followed by a made field goal less the proportion of miss streaks followed by a made field goal. Ultimately, despite the seemingly hot-handed nature of these players on the court, Daks et al. were unable to find a statistically significant overall hot hand effect for any of these players (although they were able to find a statistically significant effect in certain games) [7].

Wang and Fan implemented a simple linear regression model on modern NBA player data wherein the first two independent variables each correspond to a players' shooting percentage in consecutive games, and the dependent variable is the shooting percentage in the third game. They found that the accuracy of the model was below 40% and rejected the existence of the hot hand phenomenon [18].

McNair et al. also attempted to find evidence of the hot-hand effect in larger modern NBA datasets using the same test statistic as Miller & Sanjurjo (see above) but only with streaks of size one. Overall, after replicating previous studies, including GVT (1985), they found no evidence of a hot-hand effect in real-world NBA data. Moreover, the authors found no evidence for the notion that players on a hot streak will take more difficult shots (often referred to as a 'heat check'), although the authors acknowledge that determining patterns in shot difficulty is extremely difficult. Interestingly, the authors do not attempt a permutation test similar to that outlined in Miller and Sanjurjo (2014) [10].

Other research provides some evidence to the contrary. Ritzwoller and Romano studied permutation tests of the null hypothesis of randomness based on a similar test statistic to Daks (2017). The asymptotic distributions of these test statistics and permutation distributions were characterized under randomness, under both a general class of stationary processes and Markov chains, allowing the researchers to determine local asymptotic power. The results were then applied in an attempt to find evidence for the hot hand effect. In one of the trials, Ritzwoller and Romano were able to show that one shooter had a shooting pattern that was inconsistent with randomness. These findings indicated that a larger data set is needed to determine whether a significant deviation from randomness occurs for "streak" players. Moreover, their findings necessitate a direct test of the hot hand fallacy [16].

Similarly, Chang S.C. used NBA player tracking data from the 2015-2016 season to analyze top-players' performance with the intention of identifying any possible hot hand effect. An examination of the results led Chang to conclude that more research is needed to properly investigate the hot hand effect. Moreover, given that players participate in different numbers of games, Chang suggested that a mixed model or hierarchical model be used. [4].

The vast majority of literature examining the hot hand in basketball assumes that shot selection is independent of how hot or cold the player believes themselves to be. Bocskocsky, Ezekowitz, and Stein leverage player tracking data and shot tracking data to show that when players perceived themselves as hot, they were not only more likely to take more difficult shots, but were also more likely to be more heavily defended by the opposing team, indicating that these variables are dependent. After correcting for shot difficulty, it was estimated that a hot hand effect was present with a magnitude of 1.2 percent increase in shot accuracy for each prior field goal made.

Similarly, Pelechrinis et al. considered the streak selection bias correction implemented by Miller and Sanjurjo and applied it to real-world basketball data, while also taking shot difficulty into account. This data was obtained from the SportVU optical tracking system for both the 2013-14 and 2014-15 seasons, resulting in a data set comprising roughly 400,000 shot attempts. Pelechrinis et al. not only applied the permutation test to remove small sample bias but also created a model for shot quality. To do-so, they created a feedforward neural network with four hidden layers; ultimately, this model had an accuracy of 66% with regards to shot-make probability. Given the sequence of shots and newly generated shot difficulty vectors, Pelechrinis et al. then simulated the permutation test on the sequence of shots, and were able to find significant levels of streakiness. More specifically, of the 153 players eligible to be examined, 24 had a statistically significant streakiness effect.

1.2 Identifying a Gap in the Research

While the results from Miller and Sanjurjo were certainly compelling, the results have proven to be difficult to replicate using modern NBA datasets [7, 13]. Other approaches (which are outlined above) have generally found that streakiness is either insignificant or has little impact on field goal outcome. One commonality between these papers is that they do little to investigate hot-handedness at the individual player level. Most current literature uses the same test statistic on every player [7, 13], which does not take any player-specific effects into account at all.

In this paper, we posit that such approaches may not be robust enough to capture the effect of streakiness in shooting. Indeed, if a 'hot hand' effect exists, it should be defined such that it is relative to a player's ability to make a shot. After all, an average shooting night for an NBA superstar might be a 'hot' night for a more typical NBA player. As a result, we elected to investigate 'hot-handedness' in the context of player-specific models, each of which contains fixed effects for its respective player. Such models would allow us to differentiate standards for streakiness on a player-by-player basis, while also giving perspective on which specific players might be, on average, more likely to experience a 'hot hand' effect, unlike the pooled models used by Arkes and Bocskocsky [1, 3].

We hypothesized that this approach would allow us to gain insight into how streakiness factors into a player's field goal percentage. We also anticipate that, should a 'hot hand' effect exist, we may be able to use our results to classify players into different categories of 'streakiness.'

1.2.1 Research Questions

1. Can we create models for field goal percentage that incorporate individual fixed effects and streakiness?
2. Can we determine how significant streakiness is in predicting field goal outcome at an individual level?
3. How might we use evidence of a 'hot hand' effect, if any, to classify players as 'streaky' or 'not streaky'?

2

Methodology

2.1 Dataset

We use a data set comprised of modern NBA play-by-play data (2015-2020) from **Kaggle**, a popular site that hosts a number of data sets. The data set was scraped from **Basketball Reference**, which hosts several years of play-by-play data [17].

The play-by-play data contained a number of plays that were not initially relevant to the exploratory data analysis. As a result, all plays were removed with the exception of shots. Given that free throws were also counted as a separate category of shot, they were either re-categorized such that they fell under the normal shot type category or excluded from the data set, depending on the analysis that was being conducted. For example, when replicating other studies, we include free throws to remain consistent with the literature, but our own methods do not incorporate free throws into the analysis.

We define a streak as two or more makes in a row. Each streak is assigned an identifier between $[1, \ldots k]$, where k is the last element in the streak. This number represents the numerical position of each field goal attempt within a streak. Notably, streaks are reset on a per-game basis, such that streaks do not extend between different dates.

2.2 Exploratory Data Analysis

We first investigate different field goal percentages for each shot taken by a player given k successive makes. We generate a series of charts, wherein overall field goal percentage for the season is represented by a red line, and subsequent make probabilities after k successive makes are depicted as a line graph, wherein $k = 1$ is equivalent to the first shot after a make, $k = 2$ is first shot after two makes, and so on and so forth (see Figure 2.2.1). While a selection of studies included free throws in their calculation of field goal percentages, free throws are excluded from the calculations herein, so as to remain consistent with the way the NBA determines field goal percentage. Note that there is some variability between players' field goal percentage calculated here and players' official field goal percentage as calculated by the NBA. This difference is likely due to missing data, and appears to vary by around half a percentage point on average.

As expected from the results of previous research [8, 13], most of these players experience a decline in field goal percentage as they experience increasingly longer streaks (see Figure 2.2.1). There are, however, a few notable exceptions. James Harden, then playing for the Houston Rockets, showed significant increases in field goal percentages as streaks increased in length, with a jump in efficiency of

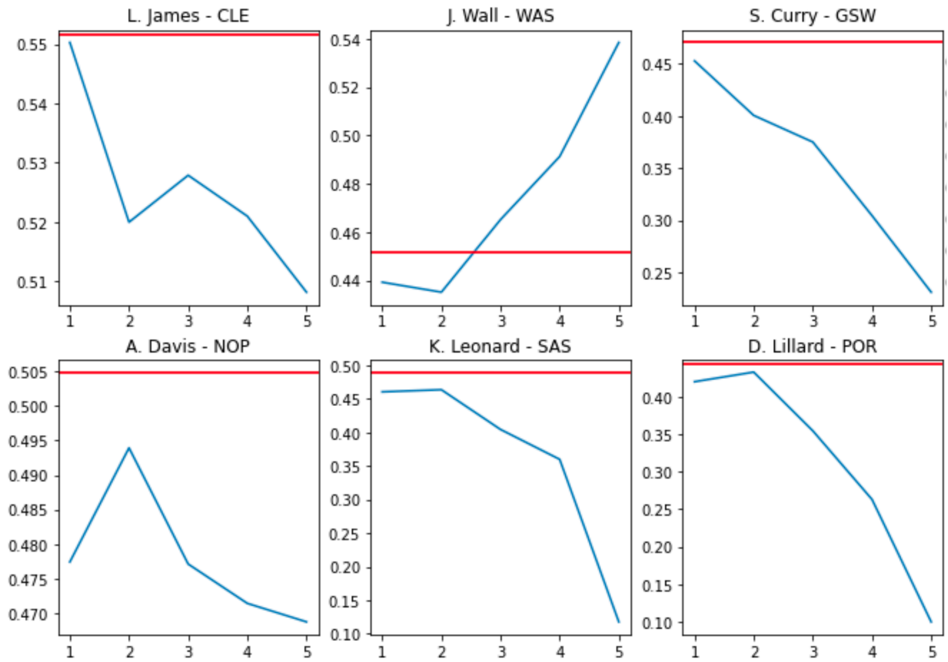

Figure 2.2.1: Shooting percentages given k previous makes for six selected high volume shooters in the 2016-2017 NBA season. Red line corresponds to overall percentage. Blue line corresponds to field goal percentage as a function of number of made field goals in a row.

$\approx 15\%$ over their average field goal percentage. Perhaps the most extreme example, however, was Klay Thompson, who appears to consistently experience higher field goal percentages, even after a single made shot (see Figure 2.2.2).

2.3 Replicating Miller and Sanjurjo

Given this newfound information with regards to increases in field goal percentage over time in certain players, we wanted to examine these players' streakiness in the context of the latest research conducted by Miller & Sanjurjo, as well as Daks, Desai, & Goldberg [7, 13]. As a reminder, Miller and Sanjurjo identified a bias in the

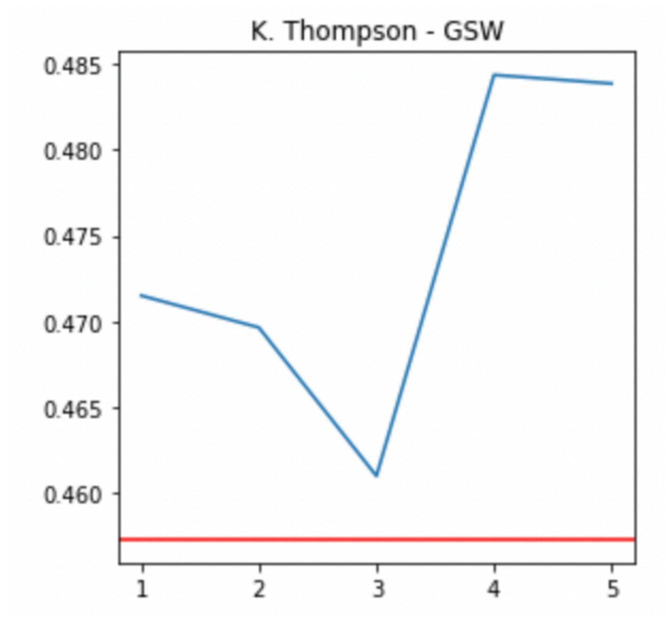

Figure 2.2.2: Field goal percentage for Klay Thompson given k previous makes. Red line corresponds to overall percentage. Blue line corresponds to field goal percentage as a function of number of made field goals in a row.

calculation of make probabilities that was directly caused by the selecting streaks from a small sample size. Daks et al. attempted to remedy this small sample size bias by performing 10,000 permutation tests on each game-long set of shooting data for each player by computing the following test statistic during each run.

$$t_k = t_{k,hit}(X) - t_{k,miss}(X).$$

Here X is a binary string of game data, k is the length of the streak, $t_{k,hit}$ is the conditional fraction of hits given k prior hits, and $t_{k,miss}$ is the conditional fraction of misses given k prior misses. For clarification, consider the following binary string:

$$111010110110111.$$

Assuming $k = 2$, we see that there are 4 instances of 11 in the above binary string. Two instances of 11 are followed by a made field goal, so $t_{k,hit} = \frac{2}{5}$. Three instances of 11 are followed by a missed field goal, so $t_{k,miss} = \frac{3}{5}$. Thus, $t_k = \frac{2}{5} - \frac{3}{5} = -\frac{1}{5}$.

This method was implemented in Python as specified by Daks et. al: the input was binary strings representing field goal attempts, and then permuted the input $10,000$ times, calculating the test statistic for each permutation. Finally, we compared the test statistic from the original string to those of the bootstrapped data to calculate the p value. The model was then run it on a subset of the top players in the NBA, as defined by shooting volume. The results herein were then confirmed to be similar to those of the original paper by Daks et. al., in that no players were identified as having a statistically significant 'hot hand' effect [7],

In Daks et. al. the researchers investigate the possible existence of a 'hot hand' effect in one of Klay Thompson's streakiest games — a famous moment in NBA history when Thompson scored 60 points on

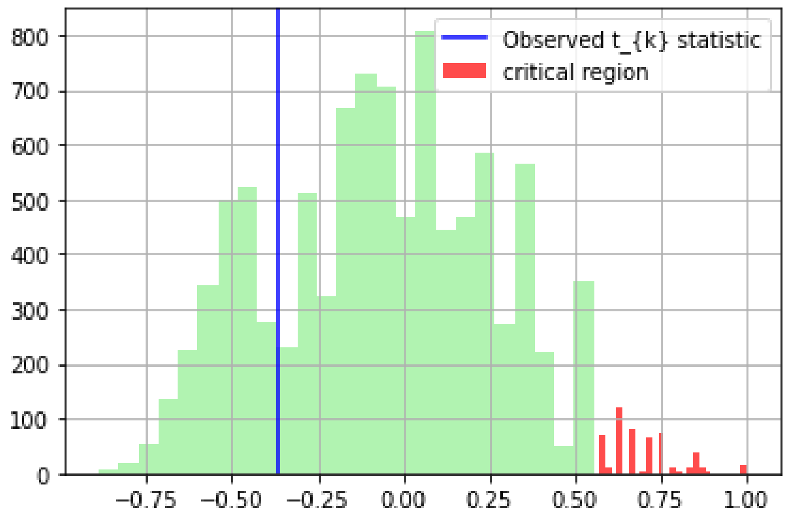

Figure 2.3.1: Histogram of test statistics for Klay Thompson's 60-point game. Red area indicates the critical region in which the test statistic would be statistically significant. Blue line corresponds to the test statistic as calculated for the original series

21-33 shooting in 29 minutes (and also with a mere 11 dribbles!). His shooting in this game can be represented by the following binary string (note that Daks et. al. elected to include free throws in their calculations)

$$110111100101111110011101111011101111101010101.$$

In order to ensure fidelity with the original paper's results, the permutation test method was conducted on these shooting data, with $k = 2$. The empirical distribution of t_k generated by our Python implementation achieved results that were equivalent to the original paper (see Figure 2.3.1).

Notably, despite the exceptional level of Thompson's playing this game, the p value generated was far from exceptional. Indeed, the result of $p = 0.8057$ indicated that the observed statistic for the game was exceeded by 80.57% of that of the $10,000$ permuted binary outcome strings, indicating that there was no hot-hand effect present in this game.

It is surprising that there was a notable lack of ‘streaky’ games found — especially when the outputted binary strings appeared to be clearly ‘streaky.’ For example, 60 point game from above represents an extraordinarily ‘hot’ shooting night, even for an incredibly talented player like Klay Thompson. Given this, we elected to dig deeper into the potential limits of this test statistic by running the method on an outcome that was intentionally created to be an obviously ‘streaky’ shooting night for a player:

$$01111111111111111111111110111111111111111111111111101111111111.$$

If there were ever to be a ‘hot’ night for a player, the above would clearly be representative of one, given that such a shooting night is statistically improbable to occur (even for the greatest NBA players of all time). Even given that Daks et. al. relied on no prior data in their methodology (or even player-specific data), one would imagine that the test statistic would have determined the existence of a ‘hot hand’ effect in the above example. Instead, however, the p value generated by the permutation test was once again insignificant, with $p = .4709$ (see Figure 2.3.2).

This result indicates that there may be some limitations to the permutation approach employed by Miller & Sanjurjo. Indeed, even if a player was found to have had an extremely ‘hot’ night similar to that of the one depicted above, the result from the permutation approach would still yield an insignificant result. As we will explore in future sections, this also indicates a need to include player-specific

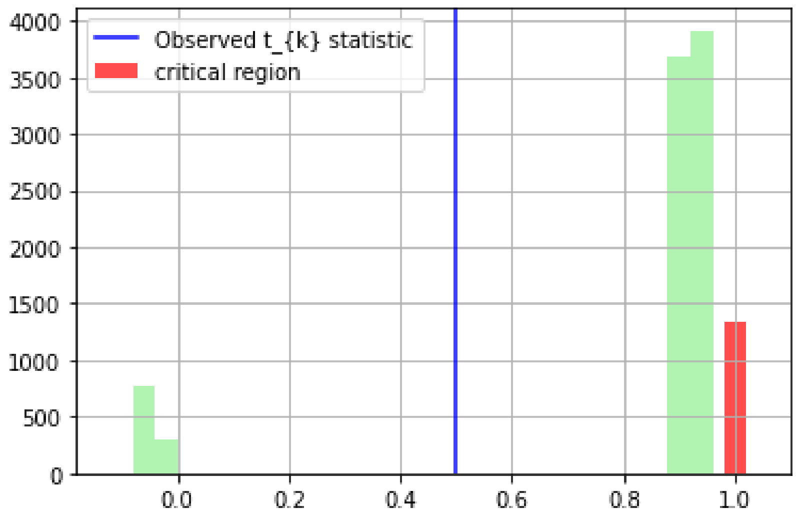

Figure 2.3.2: Histogram of test statistics for 'streaky' game. Red area indicates the critical region in which the test statistic would be statistically significant. Blue line corresponds to the test statistic as calculated for the original series

fixed effects — as indeed, this test-statistic is completely agnostic to individual players' typical performance. In other words, this example warrants an entirely different approach to the definition of streakiness, in which 'heat' is measured relative to a player's baseline performance.

2.4 Developing a Novel Approach

Most research in this field relies on the creation of novel test statistics to model streakiness in shooting. Instead, we seek to leverage supervised classification algorithms in an attempt to measure the importance of being on a hot streak in predicting player shot outcomes. More specifically, we utilized a combination of logistic regression and random forest algorithms in an attempt to quantify the importance of streakiness in predicting shot outcomes. The general approach is similar to Arkes and Bocskocsky, in that we utilize a multivariate logistic regression; that being said, our approach quantify streakiness differently. Bocskocsky and Arkes both consider whether any of the last k shots were a made field goal, and both make use of a pooled regression. Our approach utilizes indicators that describe whether only the last shot was a made field goal, and if so, what position it occupied in a streak of made field goals. Moreover, while Bocskocsky and Arkes both utilize a pooled (league-wide) logistic regression, we develop separate models for each individual player, such that we can gain greater insight into how 'hot-handedness' may materialize in individuals [1, 3].

2.4.1 Data Cleaning and Preparation

Before running logistic regressions, it was necessary to manipulate some of the variables in the data set such that the results could be properly interpreted after running our models. For instance, given that streak count is more of a categorical variable than a numerical variable, we elected to create a one-hot encoding of the variable in the dataframe. The following features were used in my models:

- **ShotDist** - The distance (in feet) from the basket from which the shot was taken
- **StreakMake_k** - A one-hot encoding for the k^{th} shot in a streak of made field goals. We only include up until the 4^{th} shot in a streak, since longer streaks likely occur too rarely to be identified as significant, and coefficients are unlikely to be reliably estimated.
- **IsHome** - A binary variable for whether the game is a home game or an away game for the player in question
- **Differential** - The difference between the player's team score and the opposing team's score at the time of the shot.

2.4.2 Variable Selection Considerations

With regard to the reasoning behind choosing these variables, my thinking was primarily aligned with creating an interpretable model. The rationale behind this decision was as follows: given that our goal was not necessarily to create the most accurate model for predicting shot outcomes, but to understand the effect of being on a streak of made field goals, it was prudent to create a simple model so as to make the impact of being on a 'hot streak' as interpretable as possible.

The selected variables and models used hereafter were selected to aid interpretability. All predictors (besides the streak indicators) that were included were those that would influence the 'pressure' on a player while engaging in a field goal attempt. For example, shot distance is a proxy for how difficult it might be to make a shot (generally speaking, the further away a shot is taken from, the harder it becomes to make the shot). 'IsHome' and 'Differential,' on the hand, reflect the mental pressure that the player might be experiencing. By including these predictors, then, we are capable of controlling for players shooting worse at away games or shooting worse when their team is winning (or losing) by a significant amount. As seen in the results, even this limited group of predictors was enough to get within reach of the best performing models for shot outcome in the present literature.

In order to ensure that models are not given any information about the outcome of the shot in question, the count of the streak is shifted down by one (if the shot was identified as the k^{th} element in a streak, it would be possible for the model to discern hidden information about the outcome of the shot). It should be noted that these models include shot data from all available years (2015-2020). This was a decision that may need to be revisited in future research, as basketball evolves at a relatively rapid pace, particularly at the professional level. For example, in 2015, roughly 26% of shots were taken from beyond the three-point line. In the current season, three-point attempts comprise 39% of field goal attempts.

index	Dist	s_1	s_2	s_3	s_4	Home	Diff
Dist	1.00	0.01	0.06	0.03	0.03	0.03	0.07
s_1	0.01	1.00	-0.22	-0.13	-0.08	-0.00	0.03
s_2	0.06	-0.22	1.00	-0.08	-0.05	0.01	0.07
s_3	0.03	-0.13	-0.08	1.00	-0.03	0.03	0.05
s_4	0.03	-0.08	-0.05	-0.03	1.00	0.02	0.05
Home	0.03	-0.00	0.01	0.03	0.02	1.00	0.11
Diff	0.07	0.03	0.07	0.05	0.05	0.11	1.00

Table 2.4.1: Correlation between features. Note that s_k is analagous to StreakMake_k

.

When analyzing the correlation between covariance parameters of the logistic regressions (both pooled and player-specific), we found the condition number was considerably higher than expected, and the determinant was extremely small. This warranted an investigation into the correlation between the features inputted into the model. Notably, there was little evidence of multicollinearity between the predictors (see figure 2.4.1). Indeed, the highest absolute correlation observed was between the streak 0 and streak 1 indicators, at $-.65$, and the magnitude of most other correlations were below .1. Given this information, we did not modify the dataset before feeding it to our models, nor did we use any particular methods to reduce multicollinearity in the dataset.

2.4.3 Pooled Logistic Regression

We began by running a logistic regression on a pooled player dataset, so as to identify any 'macro' trends across a wide sample of players.

Shooting data from the top players by shooting volume was merged and cleaned as outlined above. Then an 80%-20% train-test split was generated at random. After cleaning the data, generating streaks for each player, and combining each individual player's data into a single dataframe, there were $227,786$ observations in the train dataset. Given that we wanted to primarily investigate 'hot-handedness,' we left indicators for the 'cold' hand out of the model to isolate the hot hand effect, as we found that including the 'cold hand' indicators reduced model interpretability, in that all predictors became insignificant.

All logistic regressions were run using both the **statsmodels** and **sklearn** packages. The **statsmodels** model offered convenient access to summary statistics, and the **sklearn** model offered convenient access to other helpful **sklearn** methods. Any discrepancies between the **statsmodel** and **sklearn** models were found to be insignificant. Given that logistic regression does not require our dataset to be scaled, the dataset was not transformed beyond the steps delineated in the data cleaning and preparation section. Moreover, the logistic regressions were all run with default parameters, with the exception of the maximum number of iterations, which was increased to $20,000$ to ensure that all models converged.

2.4.4 Player-specific Logistic Regression

The natural next step in this investigation was to determine whether the outcomes found in the pooled regression were representative of performance at the individual player level. We again sampled the top 85 players by shooting volume, and ran independent logistic regressions on each player. Similarly to the pooled logistic regression, we used the same set of predictors, and created an $80\% - 20\%$ train-test split. On average, after cleaning, each of the top 40 players were associated with roughly $5,500$ observations.

Accuracy and F1 score on train and test data sets were used as a proxy for model performance. These metrics were primarily used to determine whether the models were sufficiently accurate to the point that measures of predictive significance could be considered in our analysis, as we were interested in investigating the 'hot hand' effect, rather than creating the most accurate model for shot outcomes.

2.4.5 Random Forest Hyperparameter Tuning

We conducted hyperparameter tuning on three of the parameters to the random forest model. Specifically, we attempted all possible combinations of the following with 10-fold cross-validation, and selected the parameters with the highest validation set performance. The cross-validation was conducted by using the **sklearn** GridSearchCV method.

1. **max_features**: The maximum number of features considered in a split, $set = [1, 2, 3]$

2. **max_depth**: The maximum depth of the trees in the random forest, $set = [1, 3, 5, 10, 50, None]$, where None indicates no maximum depth

3. **n_estimators**: The number of trees in the forest, $set = [10, 25, 50, 75, 100, 200]$

After tuning with cross-validation, it became clear that a **max_depth** value of 5, a **max_features** value of 2 (which is equivalent to using the square root of the number of features), and a **n_estimators** value of 100 performed the best with regards to validation performance. Given the complexity of tuning these parameters for all of the player models, and the risk of potentially overfitting, we utilize these tuned hyperparameters for both the pooled random forest and the player-specific random forest.

2.4.6 Pooled Random Forest

In the spirit of identifying how important hot-handedness might be for prediction, we elected to run a random forest on the same, simplified dataset as was used for the logistic regressions. All random forests were run using the **sklearn** package, and were run with the specifications determined in section 2.4.5. The training dataset was the same as the one specified in section 2.4.4, and had $227,786$ observations. We intended to use plots of variable importance and permutation importance to identify the relative importances of each predictor to the model. In this way, if any **StreakMake** feature appeared to have a high 'importance,' we could consider this to be evidence for the 'hot hand.' Similarly to the logistic regression, we intended to use metrics of accuracy and F1 score to assess the interpretability of the results. Since interpretability is our main focus, we did not attempt to improve the accuracy of the model far beyond a set baseline accuracy.

2.4.7 Player-specific Random Forest

Individual random forest models with the same specifications as delineated in section 2.4.5 were then run on the top players by shooting volume. Once again, we utilized the same train dataset as the individual logistic regressions. Similarly to the pooled random forest, we utilized variable importance and permutation importance to determine the relative importance of each predictor to the model and baseline metrics accuracy and F1 scores to judge the interpretability of these results.

3 Results

There is not an established best-in-class method for predicting shot outcomes. Unfortunately, although we conducted a rigorous search, we were unable to much existing information in the existing literature on modeling shot accuracy. Based on a thorough study of available methods, and the results from a Kaggle competition focused on predicting the outcome of shots in the NBA, we found that around 62% accuracy might be a reasonable bar for whether model results are interpretable. Given the significant variability in shot-taking conditions during a typical NBA game, including, but not limited to defensive pressure, shooting form, etc., it seemed unrealistic to create a model capable of achieving a significantly higher rate of accuracy. Moreover, the logistic regression and random forest models provided comparable accuracy while providing interpretable results moreso

	coef	std err	z	P> |z|	[0.025	0.975]
ShotDist	-0.0307	0.000	-90.252	0.000	-0.031	-0.030
StreakMake_1	0.1571	0.009	16.548	0.000	0.139	0.176
StreakMake_2	0.1532	0.013	11.432	0.000	0.127	0.180
StreakMake_3	0.1796	0.020	9.203	0.000	0.141	0.218
StreakMake_4	0.1890	0.029	6.488	0.000	0.132	0.246
IsHome	0.2315	0.008	30.735	0.000	0.217	0.246
Differential	-0.0007	0.000	-1.514	0.130	-0.002	0.000

Table 3.1.1: Pooled logistic regression considering league-wide data.

than other more complex algorithms considered by others.

3.1 Pooled Logistic Regression

The first regression utilizes pooled player data, and the results of the model are presented in Table 3.1.1. The accuracy of this model was around 59% on the test dataset. Notably, all streak count features have a statistically significant (at the 5% level) positive impact on the log-odds ratio, indicating the presence of a 'hot-hand' effect, at least for the pooled sample. Given that there was a possibility that only certain players may experience 'hot' shooting nights, the next step in this analysis was to explore these effects in the context of individual players.

Notably, although the streak count variables were all statistically significant, the coefficients for these features were all positive. In other words, this would actually provide support for the existence of a hot hand effect, indicating that being on a 'hot streak' actually results in an increase in the log odds of a player making a field goal.

3.2 Player-specific Logistic Regression Models

We selected the top 80 players by shooting volume and ran separate logistic regressions on the features enumerated in section 2.3.1. Interestingly, there was significant variation with regard to the potential existence of a 'hot-hand' depending on the player. The results for certain players, such as Stephen Curry or Klay Thompson found *StreakCount* to be an incredibly unimportant predictor. However, these models also had poor accuracy: the model for Curry's shooting achieved 58% accuracy on the train dataset and 55% accuracy on the test dataset.

Performance of the models with regards to accuracy did not stray far from that of the models discussed previously (see Table A.1.1 for more details). With regards to the presence of streakiness in the data, again, the majority of results did not identify being on a streak of makes to be a significant predictor. This was found to be inconsistent with the pooled results, where all streakiness predictors were found to be extremely significant ($p < 0.05$).

Of the models that did identify streakiness to be a significant predictor, the coefficients were generally positive. For comparison, the **StreakMake_k** coefficients for each player were aggregated and are shown as whisker plots in Figure 3.2.1. The IQR of **StreakMake_k** coefficients, which ranged from 0.0 to ≈ 0.25 for all streak indicators and is greater than zero for all streak counts — only the bottom quartile is less than zero. Interestingly, the coefficients were relatively steady across streak indicators, with only a slight downward tick in the coefficient from the first streak indicator to the second, followed by an upward tick from the second streak indicator to the third. As expected, the median values of the make streak indicator coefficients are aligned with the coefficients we found in the pooled logistic

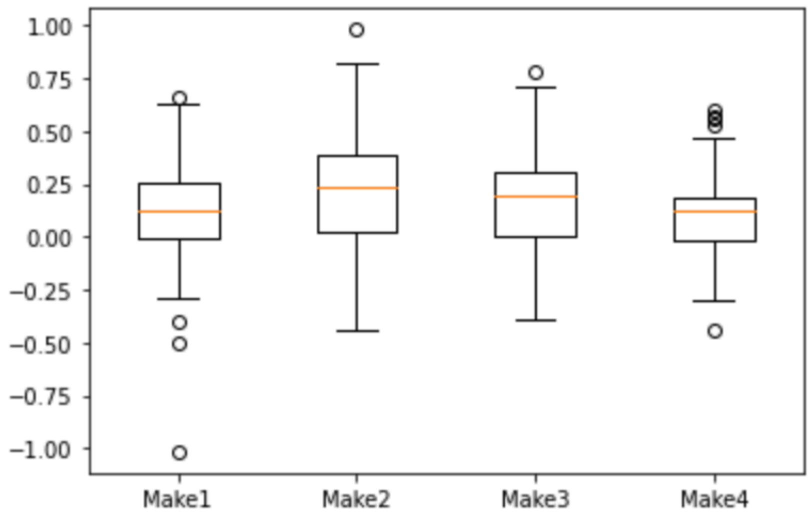

Figure 3.2.1: Boxplot of coefficients for makestreak indicators. Value of coefficient is depicted on the y axis. Results are aggregated from separate logistic regressions for the top 80 players by shooting volume.

regression.

As seen in Figure 3.2.2, the statistical significance of the streak indicators also experienced substantial variance between both players and indicators. Indeed, the median significance level for all predictors is well over $p = 0.05$. Of interest is the fact that the p values for the second indicator, **StreakMake_2**, are generally lower than that of the first indicator. One explanation may be that players that often make it to a streak of two shots in a row are more likely to be 'hot players' than those who only make it to a streak of one shot in a row. The statistical significance of each predictor deviated substantially from the results we received from the pooled logistic regression. In the pooled logistic regression, all streak indicators were statistically significant at the 0.05 level, whereas a majority of player-specific models found the predictors to be insignificant.

One possible explanation for the difference in p-value between the

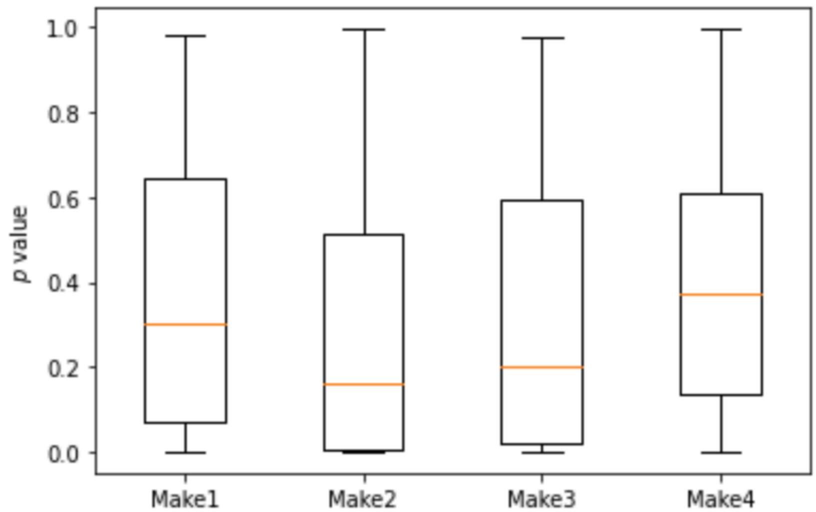

Figure 3.2.2: Boxplot of p values for makestreak indicators. Value of coefficient is depicted on the y axis. Results are aggregated from separate logistic regressions for the top 80 players by shooting volume.

pooled regression and the player-specific regressions is that the player-specific regressions simply do not have enough statistical power. In the literature, several papers, including Arkes (2010), indicated that a small sample size might negatively influence our ability to detect a 'hot hand' effect. Given that our sample sizes of, on average $5,500$ shots for individual players, is considerably smaller than that of the sample sizes used in the pooled logistic regressions in Arkes (2010) and Boc et. al., it is possible that we lack the requisite statistical power to consistently detect 'hot handedness' [1, 3].

There were still some models for players that indicated possible hot-handedness. For example, the model for Rudy Gobert (see Table 3.2.1) found streak indicators to be extremely significant predictors (with positive coefficients), indicating a potential hot hand effect.

Coefficients for the top 30 players by shooting volume are depicted in Table 3.2.2, wherein green text color indicates that the predictor

	coef	std err	z	P> |z|	[0.025	0.975]
ShotDist	-0.2148	0.022	-9.675	0.000	-0.258	-0.171
StreakMake_1	0.8256	0.108	7.639	0.000	0.614	1.037
StreakMake_2	0.7669	0.129	5.949	0.000	0.514	1.020
StreakMake_3	0.5524	0.161	3.431	0.001	0.237	0.868
StreakMake_4	0.9822	0.223	4.399	0.000	0.545	1.420
IsHome	0.8336	0.088	9.489	0.000	0.661	1.006
Differential	0.0032	0.005	0.665	0.506	-0.006	0.013

Table 3.2.1: Logistic regression on Rudy Gobert's shooting data

was statistically significant in the model, and red text color indicates that the predictor was not statistically significant in the model. Generally, when coefficients were negative, they did not tend to be statistically significant. Notably, players with statistically significant results tended to be similar in terms of playstyle, in that they generally tend to take shots closer to the basket. See the results for Giannis Antetokounmpo, Karl Anthony Towns, Anthony Davis, Andre Drummond, and Rudy Gobert (all of whom are known for playstyles that involve attacking the rim) in Table 3.2.2 as an example.

	Player	Streak 1	Streak 2	Streak 3	Streak 4
0	J. Harden - HOU	0.0405	0.1019	0.0403	−0.0052
1	G. Antetokounmpo - MIL	0.2949	0.3845	0.4554	0.1761
2	D. Lillard - POR	0.0354	0.1063	0.1374	0.3491
3	B. Beal - WAS	0.2567	0.1233	0.2518	−0.1545
4	R. Westbrook - OKC	0.1315	0.0262	0.3426	0.1741
5	C. McCollum - POR	0.1753	−0.0517	−0.0094	0.4262
6	D. Booker - PHO	0.1865	0.1001	−0.0084	−0.1525
7	L. Aldridge - SAS	0.237	0.2348	0.1838	0.4726
8	K. Durant - GSW	0.3498	0.0759	0.4729	0.2485
9	A. Wiggins - MIN	0.0151	0.1737	−0.0106	0.2562
10	K. Towns - MIN	0.3533	0.2252	0.4093	0.4196
11	S. Curry - GSW	0.1846	−0.0697	−0.272	−0.1551
12	A. Davis - NOP	0.2748	0.399	0.5673	0.4825
13	K. Walker - CHO	0.1536	0.1316	0.1869	0.382
14	D. Mitchell - UTA	0.0054	0.0474	0.2006	0.1578
15	J. Embiid - PHI	0.1592	0.2659	−0.0226	0.1297
16	K. Thompson - GSW	0.0745	0.1829	0.3025	0.2026
17	L. James - CLE	0.4258	0.4382	0.516	0.2076
18	J. Holiday - NOP	0.2163	0.2546	0.2158	0.0884
19	K. Lowry - TOR	0.1095	0.0074	−0.2243	−0.0729
20	K. Middleton - MIL	−0.0097	−0.0146	0.0641	0.2187
21	A. Drummond - DET	0.4527	0.4059	0.4827	0.5826
22	L. Williams - LAC	0.0188	0.1334	−0.1189	0.4195
23	J. Murray - DEN	0.0286	0.1176	0.229	0.228
24	D. DeRozan - TOR	0.2158	0.1808	0.162	0.2878
25	E. Gordon - HOU	−0.0004	−0.0798	0.1551	−0.1338
26	E. Fournier - ORL	0.1491	0.0363	−0.0286	0.1567
27	A. Gordon - ORL	0.1259	0.3039	0.133	0.5278
28	S. Dinwiddie - BRK	−0.0726	−0.0063	0.2433	−0.2051
29	R. Gobert - UTA	0.8256	0.7669	0.5524	0.9822

Table 3.2.2: Streak indicator coefficients for the player-specific logistic regression. Green indicates that the coefficient was significant in its respective model, and red indicates that the coefficient was not significant in its respective model.

3.3 Pooled Random Forest

The random forest achieved 62.5% accuracy on the pooled test dataset after hyperparameter tuning (notably, this was better than the performance achieved by the logistic regression). The F1 score is quite low, however, at just 48.5. Unlike logistic regression, the plot of variable importances indicated that the hot-hand indicators that we generated were not important relative to the other predictors in the model (see Figure 3.3.1).

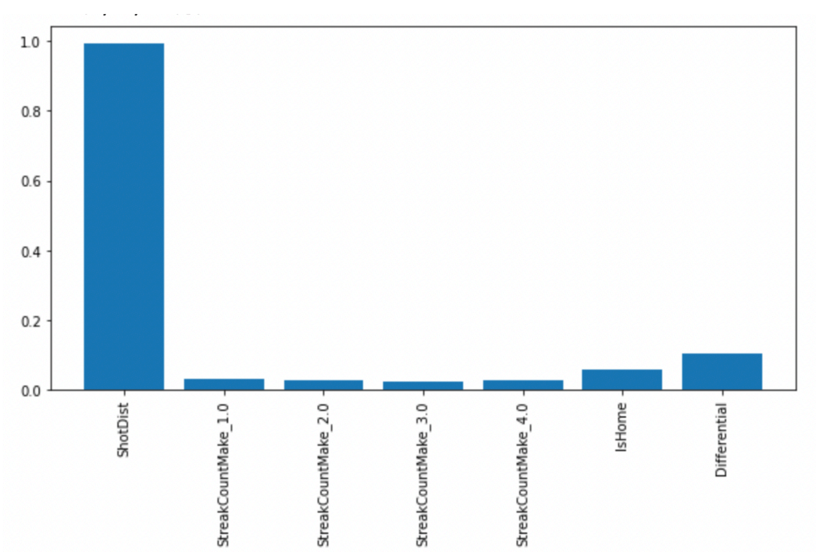

Figure 3.3.1: Feature importance for pooled random forest model. Length of bar corresponds to importance.

Of the seven predictors entered into the model, shot distance was easily the most predictive, to the point where all other predictors were effectively dominated by the importance of shot distance in making splits. This was consistent with the player-specific logistic regression results. The next most important predictor was 'differential', giving credence to our hypothesis that 'pressure-based' predictors have an effect on player shot outcome. In the case of 'differential,' this indicates that a player's team outscoring (or being outscored) by the opposing might have an effect on their chance of making a field goal. Interestingly, whereas the logistic regressions found scoring differential to be a consistently insignificant predictor, the random forest actually found it to be more significant than any of the streak predictors.

3.4 Player-specific Random Forest

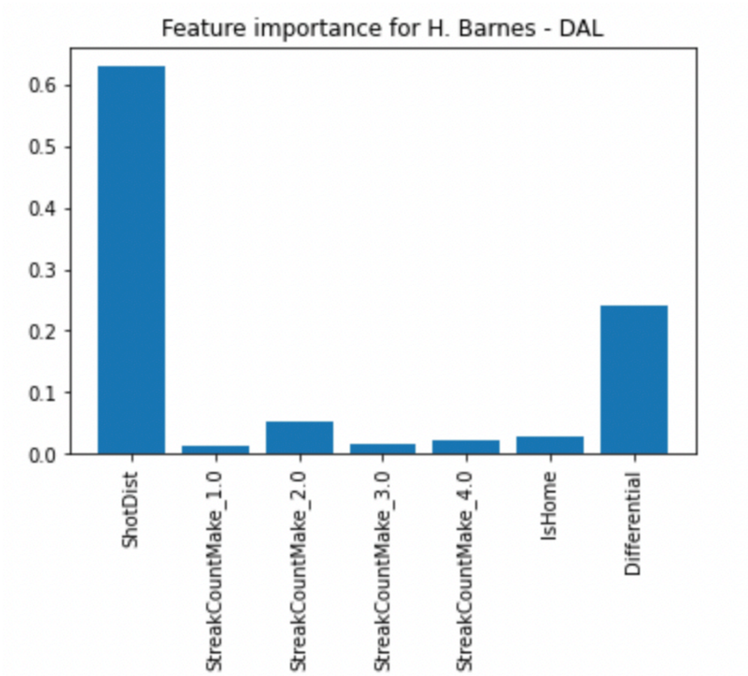

Figure 3.4.1: Feature importance for Harrison Barnes. Length of bar corresponds to importance.

As a whole, player-specific models tended to perform in-line with expectations of accuracy, with test performance generally straying by no more than a few percentage points in either direction when compared to the performance of the pooled random forest. F1 scores were extremely variable, ranging from as low as 30 to as high as 80 (see Table A.1.2).

Models for players who attempted more shots in the paint (e.g., centers or certain power forwards), also tended to be more accurate — the models for Giannis Antetokounmpo and Rudy Gobert both scored over 70% accuracy on the test dataset (which handily beats the benchmark performance we set in the introduction to the results section. With regard to the hot hand effect, similarly to the pooled random forest, most models indicated that the streakiness indicators were not important predictors relative to the others (see Figure 3.4.1 for a specific example).

Overall, we again see shot distance identified as a substantially more important predictor, followed by point differential (which varies substantially in importance by model), and finally the make streak indicators as well as the indicator for whether the respective player's team is home or away. As seen in Table 3.4.1, when we standardize importance to shot distance (which was the most important predictor for all players), streak indicators are at most 5% of the importance of shot distance, whereas differential is as much as 54% of the importance of shot distance. This is a stark difference from the results of the player-specific logistic regressions, which rarely identify differential as a significant predictor.

Player	S_1	S_2	S_3	S_4	Home	Diff
J. Harden - HOU	0.02	0.01	0.01	0.01	0.03	0.20
G. Antetokounmpo - MIL	0.01	0.01	0.00	0.01	0.01	0.08
D. Lillard - POR	0.02	0.02	0.01	0.01	0.03	0.28
B. Beal - WAS	0.02	0.02	0.01	0.02	0.03	0.23
R. Westbrook - OKC	0.02	0.02	0.01	0.01	0.03	0.21
C. McCollum - POR	0.05	0.04	0.03	0.03	0.06	0.51
D. Booker - PHO	0.02	0.02	0.02	0.02	0.03	0.32
L. Aldridge - SAS	0.02	0.01	0.01	0.01	0.03	0.21
K. Durant - GSW	0.01	0.03	0.01	0.01	0.03	0.20
A. Wiggins - MIN	0.03	0.01	0.02	0.01	0.02	0.18
K. Towns - MIN	0.01	0.02	0.01	0.01	0.02	0.21
S. Curry - GSW	0.02	0.04	0.06	0.02	0.05	0.32
A. Davis - NOP	0.02	0.01	0.01	0.01	0.03	0.15
K. Walker - CHO	0.03	0.03	0.03	0.03	0.04	0.36
D. Mitchell - UTA	0.02	0.03	0.02	0.01	0.06	0.30
J. Embiid - PHI	0.01	0.01	0.02	0.01	0.03	0.20
K. Thompson - GSW	0.02	0.02	0.03	0.02	0.05	0.38
L. James - CLE	0.01	0.01	0.01	0.01	0.02	0.13
J. Holiday - NOP	0.01	0.01	0.01	0.01	0.03	0.21
K. Lowry - TOR	0.02	0.02	0.02	0.02	0.04	0.30
K. Middleton - MIL	0.05	0.04	0.03	0.03	0.07	0.54
A. Drummond - DET	0.02	0.01	0.01	0.01	0.03	0.21
L. Williams - LAC	0.03	0.04	0.04	0.03	0.07	0.47
J. Murray - DEN	0.03	0.03	0.02	0.02	0.06	0.34
D. DeRozan - TOR	0.02	0.02	0.02	0.02	0.04	0.32
E. Gordon - HOU	0.03	0.03	0.02	0.01	0.04	0.33
E. Fournier - ORL	0.02	0.02	0.02	0.01	0.04	0.37
A. Gordon - ORL	0.01	0.01	0.01	0.01	0.02	0.20
S. Dinwiddie - BRK	0.03	0.02	0.01	0.02	0.04	0.26
R. Gobert - UTA	0.02	0.01	0.01	0.02	0.05	0.22

Table 3.4.1: Player-specific random forest feature importances. All values are standardized to the feature importance of shot distance (and shot distance was subsequently dropped from this table).

4

Discussion

4.1 Conclusions and Implications

This study yielded a number of insights into the potential existence of a hot-hand effect in basketball. Previous studies that found evidence for the 'hot hand' have predominantly relied on developing test statistics based off conditional probabilities of made or missed field goals. While other studies have also utilized regression techniques, our methodology and data representation provide a unique approach to investigating the existence of streakiness in shooting. Indeed, unlike other studies, rather than investigating the existence of the hot hand independently, we examine the degree to which streakiness may play a role in predicting shot outcomes and identify player-specific differences in hot-handedness.

Our pooled logistic regression indicated that, across the top 40 players in the league, being on a streak not only was a significant predictor of field goal outcome, but also increased the log odds of the player making this shot — an indication that a ‘hot hand’ may indeed exist. Player-specific logistic regressions gave insight into the existence of this effect at an individual level. As expected, the significance and direction of ‘hot handedness’ as a predictor varied considerably depending on the player — for certain players, the effect was significantly more pronounced, whereas for others, it was completely insignificant. Interestingly, players with shot profiles in the paint tended to have a profile more consistent with streaky shooting. Given that shots in the paint are generally more likely to go in than those from beyond the arc, this trend appears to seem reasonable.

The random forest classifiers provide an entirely different perspective than the logistic regressions. The pooled random forest provided no indication that streakiness was important to predicting shot outcomes relative to shot distance and differential. Upon plotting the feature importances (and the permutation importances), we found that the models determined our streak indicators to be relatively unimportant to the model.

One particularly interesting case was that of Rudy Gobert, whose model (among many other centers and power forwards) found all four make streak predictors to be extremely significant in his logistic regression ($p \leq 0.001$, see figure 4.1.1). As with other players, however, the random forest model found the streak predictors to be extremely unimportant (see 4.1.1. More interestingly, the random forest for Gobert's shooting data found *differential* to be a relatively important predictor, even when the logistic regression on his shooting data found differential to be insignificant $p \approx .5$. Much like the ongoing research on this topic, we are left with an ambiguous result, wherein one model indicates a streakiness effect, and another indicates that streakiness is unimportant.

	coef	std err	z	P> |z|	[0.025	0.975]
ShotDist	-0.2148	0.022	-9.675	0.000	-0.258	-0.171
StreakMake_1	0.8256	0.108	7.639	0.000	0.614	1.037
StreakMake_2	0.7669	0.129	5.949	0.000	0.514	1.020
StreakMake_3	0.5524	0.161	3.431	0.001	0.237	0.868
StreakMake_4	0.9822	0.223	4.399	0.000	0.545	1.420
IsHome	0.8336	0.088	9.489	0.000	0.661	1.006
Differential	0.0032	0.005	0.665	0.506	-0.006	0.013

Table 4.1.1: Logistic regression on Rudy Gobert's shooting data

A possible explanation for the discrepancies between the random forest model results and that of the logistic regressions lies with the importance of *ShotDist*. Given that shot distance so regularly influences the degree of difficulty of the shot, it is possible that it 'explains away' any effects from being on a 'hot streak,' at least in the random forest models. Similarly to Boc et. al., we attempted to investigate the interaction between make streaks and shot distance in the context of predicting shot outcomes but found that this substantially reduced model interpretability (and potentially as a result of our relatively small datasets for individual players).

4.2 Comparison to Previous Works

As with other studies, the nature of the 'hot hand effect' is once again ambiguous here. Although our logistic regressions would seem to indicate the existence of such an effect (at least in certain individuals' playing styles), the random forest indicates the opposite. That being said, our methods bring a valuable new approach and perspective to

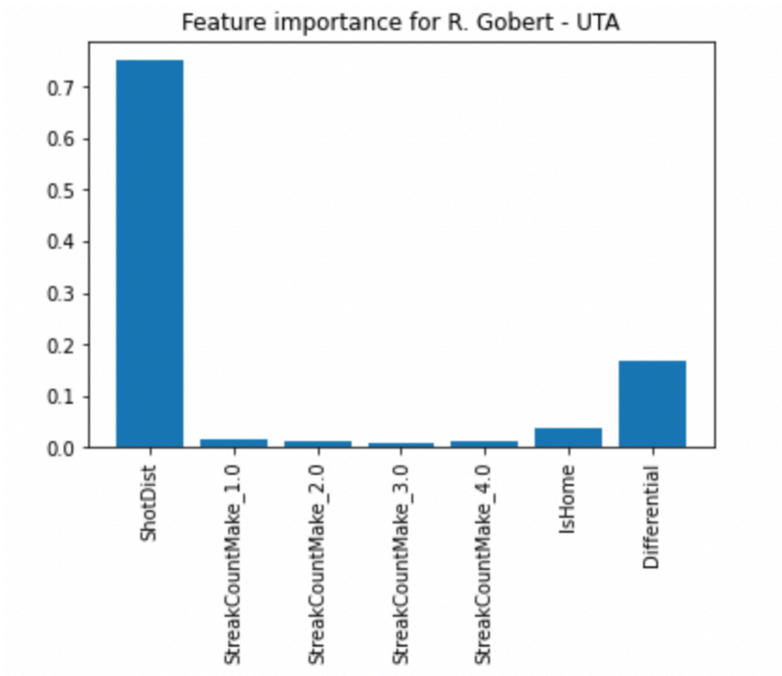

Figure 4.1.1: Feature importance for Rudy Gobert. Length of bar corresponds to higher importance.

this problem.

As we mentioned earlier in this paper, some of the most significant studies on hot-handedness have utilized a difference in conditional probabilities (on both make and miss streaks), in order to determine the existence of a streakiness effect in shooting [7, 13]. As we show in Chapter 2, however, this approach has significant limitations. Recalling our original example of the limitations of this method, we ran the bootstrapping and test statistic on the following shooting night:

$$0111111111111111111110111111111111111111111111101111111111.$$

This should be an unquestionably 'hot shooting night,' yet Miller and Sanjurjo's test statistic actually identifies streakiness as insignificant in this series of shots, with a p value of .4709 (see Figure 2.3.2, calling into question the way researchers are setting up this problem and the manner in which current research defines streakiness in shooting in general.

Perhaps one issue with the approach of existing literature (including other studies that have more similar methods to the one we utilize here, such as Bocskocsky et. al. [3]), is the absence of player-specific effects. Indeed, while Miller and Sanjurjo do investigate their test statistic at an individual player level, they do not include any information about that player's abilities, and instead run the same test statistic on each player's data. In this paper, we assume that streakiness is unique to each individual player —— and that a 'hot hand' effect should be relative to each individual player's ability to make a shot. Thus, standards for streakiness should be different from player to player (hence our use of player-specific logistic regression). For example, a regular shooting night for a player like

Stephen Curry might be considered an extremely 'hot' shooting night for another player.

4.3 Recommendations for Future Work

Our work in this paper clearly indicates the need for future research. For one, we elected to represent 'hot streaks' as indicators of having made the k^{th} previous shot in a streak, whereas previous papers have utilized a measure of heat based on the proportion of previously made field goals [3]. Future projects should consider different methods of representing streakiness and their respective advantages and disadvantages.

We also recognize the limits of sample size in our models, particularly with regard to player-specific regression. Perhaps future studies might consider employing a bootstrapping approach in order to improve sample size, and avoid running into the issue presented by the 'Truth in the Law of Small Numbers' as defined by Miller & Sanjurjo [13].

We also sought to create the most interpretable model results possible, rather than creating more complicated (but potentially more accurate) models. One could certainly consider exploring this problem with a more complicated model (such as a neural network, gradient boosting, etc.).

Exploring hot-handedness in the context of shot difficulty was also of interest to us, but unfortunately, our data set was limited. Ultimately, we were only able to use shot distance, team point differential, and whether the team was home or away, as proxies for shot difficulty. It could prove fruitful to build models that include more in-depth information on shot difficulty, similarly to Bocskocsksy et. al. [3]. Optical player tracking data (such as the datasets generated by SportsVU), could prove invaluable to future research.

Such data could also be used to explore the existence of ‘heat checks,’ or the idea that players tend to take increasingly difficult shots after making consecutive field goals.

4.4 Conclusion

Although this paper does not provide a definitive answer as to the existence of a ‘hot hand’ effect, it does provide limited evidence towards its existence. More importantly, our results demonstrate that should a ‘hot hand’ exist, it will vary from player to player to the extent that some players may experience extreme streakiness, and others may not experience it at all. Our results also give credence to the idea that, should future researchers endeavor to build a model for shot outcomes, they can consider using streakiness as a way to further tailor such a model to individual players. Such insight would also be of great interest to NBA franchises and their increasing reliance on data-driven methods for player evaluation, by allowing coaches to make more informed decisions about prioritizing players at certain points during games.

A
Appendix

A.1 Description of Figures

1. Accuracy and F1 score for player-specific logistic regression: A.1.1
2. Accuracy and F1 score for player-specific random forest classification: A.1.2

Player	Train	Test	Train F1	Test F1
J. Harden - HOU	0.606486	0.600837	0.522014	0.534959
G. Antetokounmpo - MIL	0.657099	0.645075	0.715098	0.704142
D. Lillard - POR	0.586305	0.583607	0.480020	0.466387
B. Beal - WAS	0.601593	0.600165	0.523798	0.538535
R. Westbrook - OKC	0.622621	0.636121	0.526257	0.560687
C. McCollum - POR	0.557374	0.529958	0.426942	0.383167
D. Booker - PHO	0.588265	0.598990	0.486334	0.492976
L. Aldridge - SAS	0.590731	0.588004	0.586779	0.586377
K. Durant - GSW	0.593882	0.566596	0.620690	0.587525
A. Wiggins - MIN	0.626313	0.629667	0.550271	0.541823
K. Towns - MIN	0.598674	0.599787	0.655117	0.661858
S. Curry - GSW	0.581638	0.569231	0.488654	0.467005
A. Davis - NOP	0.624144	0.621714	0.653330	0.656995
K. Walker - CHO	0.583288	0.580715	0.382094	0.394366
D. Mitchell - UTA	0.596325	0.619306	0.458303	0.504937
J. Embiid - PHI	0.609552	0.634409	0.600068	0.631436
K. Thompson - GSW	0.559940	0.551896	0.419260	0.400534
L. James - CLE	0.644395	0.634518	0.686050	0.694915
J. Holiday - NOP	0.598621	0.605759	0.590487	0.607930
K. Lowry - TOR	0.595398	0.608974	0.451948	0.483926
K. Middleton - MIL	0.561162	0.553922	0.440982	0.401316
A. Drummond - DET	0.632323	0.609655	0.700533	0.683799
L. Williams - LAC	0.602881	0.590659	0.380086	0.322727
J. Murray - DEN	0.587417	0.592233	0.457870	0.452769
D. DeRozan - TOR	0.595409	0.593391	0.518360	0.499115
E. Gordon - HOU	0.623642	0.615187	0.461116	0.412574
E. Fournier - ORL	0.594265	0.592348	0.507008	0.516432
A. Gordon - ORL	0.636996	0.642061	0.580386	0.587480
S. Dinwiddie - BRK	0.637429	0.600000	0.531593	0.489627
R. Gobert - UTA	0.713860	0.711155	0.814791	0.812903
J. Tatum - BOS	0.594240	0.597668	0.516717	0.515789
P. George - OKC	0.602692	0.602465	0.428334	0.479839
K. Love - CLE	0.589774	0.584375	0.351635	0.348039
B. Hield - SAC	0.566711	0.582888	0.364170	0.417910
J. Brown - BOS	0.606095	0.569678	0.593234	0.544571

Table A.1.1: Accuracy and F1 score for player-specific logistic regression.

Player	Train	Test	Train F1	Test F1
J. Harden - HOU	0.634582	0.635729	0.447706	0.452830
G. Antetokounmpo - MIL	0.703025	0.718722	0.707926	0.724108
D. Lillard - POR	0.615447	0.614754	0.374544	0.352617
B. Beal - WAS	0.623647	0.608409	0.447045	0.460840
R. Westbrook - OKC	0.657592	0.645018	0.483817	0.508015
C. McCollum - POR	0.586952	0.560338	0.315607	0.241630
D. Booker - PHO	0.615966	0.616162	0.404994	0.389068
L. Aldridge - SAS	0.621628	0.614553	0.505607	0.492228
K. Durant - GSW	0.632120	0.578224	0.589585	0.522156
A. Wiggins - MIN	0.649825	0.655903	0.485294	0.484115
K. Towns - MIN	0.632626	0.625400	0.601897	0.602492
S. Curry - GSW	0.593082	0.584615	0.413490	0.392804
A. Davis - NOP	0.659817	0.643429	0.615484	0.609023
K. Walker - CHO	0.618868	0.583965	0.345370	0.316726
D. Mitchell - UTA	0.632262	0.623644	0.381645	0.392294
J. Embiid - PHI	0.656313	0.655914	0.550066	0.549296
K. Thompson - GSW	0.598163	0.562874	0.364849	0.284314
L. James - CLE	0.685877	0.676396	0.666891	0.676806
J. Holiday - NOP	0.630897	0.638981	0.513100	0.542135
K. Lowry - TOR	0.635027	0.624359	0.401468	0.390852
K. Middleton - MIL	0.607339	0.566176	0.400000	0.319231
A. Drummond - DET	0.665403	0.671724	0.699288	0.697201
L. Williams - LAC	0.628944	0.603022	0.390766	0.341686
J. Murray - DEN	0.628252	0.597087	0.380232	0.299578
D. DeRozan - TOR	0.623745	0.616379	0.460113	0.402685
E. Gordon - HOU	0.649840	0.643501	0.423764	0.369021
E. Fournier - ORL	0.620303	0.629288	0.414039	0.425358
A. Gordon - ORL	0.682197	0.664345	0.572897	0.546139
S. Dinwiddie - BRK	0.671382	0.613008	0.519219	0.454128
R. Gobert - UTA	0.731744	0.725100	0.822017	0.818421
J. Tatum - BOS	0.631425	0.609329	0.488101	0.429787
P. George - OKC	0.638846	0.622496	0.408318	0.426230
K. Love - CLE	0.619048	0.621875	0.351926	0.349462
B. Hield - SAC	0.612742	0.590909	0.336192	0.337662
J. Brown - BOS	0.650667	0.592649	0.561454	0.472222

Table A.1.2: Accuracy and F1 score for player-specific random forest classification.

References

[1] Jeremy Arkes. Revisiting the hot hand theory with free throw data in a multivariate framework. *Journal of Quantitative Analysis in Sports*, 6, 1 2010.

[2] Rob Arthur and Greg Mathews. Baseball's 'hot hand' is real | fivethirtyeight, 2017.

[3] Andrew Bocskocsky, J. Ezekowitz, and Carolyn Stein. The hot hand : A new approach to an old "fallacy". 2014.

[4] Shun Chuan Chang. Capability and opportunity in hot shooting performance: Evidence from top-scoring nba leaders. *PLOS ONE*, 13:e0179154, 2 2018.

[5] Peter Coy. The day stephen curry made 105 three point shots in a row. *Bloomberg.com*.

[6] Peter Csapo, Simcha Avugos, Markus Raab, and Michael Bar-Eli. The effect of perceived streakiness on the shot-taking behaviour of basketball players. *European journal of sport science*, 15:647–654, 10 2015.

[7] Alon Daks, Nishant Desai, and Lisa R. Goldberg. Do the golden state warriors have hot hands? *SSRN Electronic Journal*, 11 2017.

[8] Thomas Gilovich, Robert Vallone, and Amos Tversky. The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, 17:295–314, 1985.

[9] Robert Lantis and Erik Nesson. Hot shots: An analysis of the "hot hand" in nba field goal and free throw shooting. *Journal of Sports Economics*, 22:639–677, 8 2021.

[10] Brian McNair, Eric Margolin, Michael Law, and Ya'acov Ritov. The hot hand and its effect on the nba. 10 2020.

[11] Joshua Miller. Momentum isn't magic-vindicating the hot hand with the mathematics of streaks. *Scientific American*, Mar 2018.

[12] Joshua B. Miller and Adam Sanjurjo. Surprised by the hot hand fallacy? a truth in the law of small numbers. *Econometrica*, 86:2019–2047, 11 2018.

[13] Joshua Benjamin Miller and Adam Sanjurjo. Is it a fallacy to believe in the hot hand in the nba three-point contest? *SSRN Electronic Journal*, 6 2015.

[14] Joshua Benjamin Miller and Adam Sanjurjo. A visible (hot) hand? expert players bet on the hot hand and win. *SSRN Electronic Journal*, 12 2017.

[15] Joshua Benjamin Miller and Adam Sanjurjo. A cold shower for the hot hand fallacy: Robust evidence that belief in the hot hand is justified. *SSRN Electronic Journal*, 8 2019.

[16] David M Ritzwoller and Joseph P Romano. Uncertainty in the hot hand fallacy: Detecting streaky alternatives to random bernoulli sequences. *The Review of Economic Studies*, 7 2021.

[17] Schmadamco. Nba play-by-play data 2015-2021. https://www.kaggle.com/datasets/schmadam97/nba-playbyplay-data-20182019.

[18] Jingru Wang and Qishi Fan. Application of machine learning on nba data sets. *Journal of Physics: Conference Series*, 1802:032036, 3 2021.

[19] Joseph Zucker. Klay thompson sets nba record with 37 points in quarter: Stat line and reaction. *Bleacher Report*, Oct 2017.

[20] Marius Ötting and Groll Andreas. A regularized hidden markov model for analyzing the 'hot shoe' in football:. *https://doi.org/10.1177/1471082X211008014*, 5 2021.