MWSUG 2017 - Paper AA13

How Can an NBA Player Be Clutch?: A Logistic Regression Analysis

Logan Edmonds, Oklahoma State University, Stillwater, OK

ABSTRACT

Many NBA players are known as clutch shooters that put fear in the opposing team as the clock is winding down. Michael Jordan is remembered as much for his game ending shots as the high-flying dunks. We know that these players can make the shot, but are there certain situations that contribute to the likelihood the game winner will go in? Using PROC LOGISTIC, this paper determines the key components of made shots during the crucial last two minutes of an NBA game. Using shot log data from the 2014-2015 NBA season, over 120,000 shots were filtered to those occurring in the last two minutes of regulation or overtime. Many things can affect whether a player will make these high pressure shots. Specifically, the effects of home court, back-to-back game fatigue, shot distance and dribble time before the shot are considered as possible predictors of clutch shooting. Assessment of only final game shots includes discussion of how teams could potentially use these variables in designing end of game plays. Finally, this analysis seeks a quantitative analysis of how players considered by pundits as 'clutch' shooters live up to their billing.

INTRODUCTION

The data science site Kaggle recently released data (Kaggle 2016) about each individual shot taken during the 2014-2015 NBA season using information from the STATS SportVu® tracking system. The data set used in this analysis contains data on shots taken during individual games up until March 3rd, 2015. This particular season concluded on April 4th, 2015. In total this data set contains 903 unique NBA games, approximately 73% of the 2014-2015 season. For a full NBA season of 82 games, we would expect a total of 1,230 games during the regular season, not including playoff games.

128,069 individual shots appear in the data from 281 players. Based upon the standard NBA roster size of 15 players for each of the 30 NBA teams, we expect at least 450 players to appear as shooters (even the worst players get court time). 470 players appear as defenders, which is possible given that teams waive/drop players and pick up NBA Developmental League (equivalent to MLB minor leagues) players midseason. In total 189 players do not appear as shooters who likely should. Many of these missing shooters are minor players who likely did not have many shot attempts during the season, but some notable names are missing. In particular, former NBA MVP Kevin Durant is not included as a shooter.

In addition to shot data, data pulled from NBA Reference is included for information on player heights and personal information (birthplace, college, position, etc.).

Previous analyses (Cen, et al., 2015, Zhang 2016) have used various modeling techniques to predict NBA shot outcomes, but none of these focused on how players shoot during clutch situations. This paper examines how players shoot during these situations and in what ways these situations differ from non-clutch scenarios.

Binary shot outcomes are modeled for both clutch and non-clutch situations to compare accuracy of predictions and to determine if significant predictors differ between the shot types. There are multiple methods for performing binary classification, but for the sake of simplicity of interpretation and previous results of modeling NBA shots I have chosen to use a logistic regression classification model.

I have defined ‘clutch’ as situations in which a shot was taken in the final two minutes of regulation or overtime in a game decided by 5 or fewer points. Of the 902 games included in the data set, 252 (27.9%) qualify as ‘close’ based upon the above criteria of a final margin of 5 or less. An unfortunate artifact of close games in the NBA is that teams will often foul late in games to try and extend the game when the outcome is still in doubt. Many games that are close in the final minute may end up with inflated final margins due to the winning team shooting a large number of free throws at the end of regulation. Inversely, blowout games may have final margins of 5 or fewer when the losing team makes multiple meaningless 3 pointers in the final minute. My definition for clutch shooting is imperfect, but should capture a majority of the high pressure situations encountered during the 2014-2015 season.

Table 1 introduces variables in the NBA Shots and player information data sets that were considered as potential modeling predictors, as well as predictors derived from these variables.

Possible Predictor Variable Name	Description
*catch_shoot	indicates whether player dribbled prior to shot
close_def_dist	distance in feet of closest defender prior to shot
defender_height	defender height in inches
Dribbles	number of dribbles taken prior to shot
final_margin	final point differential of shooter's team (negative for losing team)
*height_diff	shooter height – defender height (positive values indicate taller shooter)
Location	shooter is at home or away venue
period_sec	number of seconds remaining in period
Pos	position of shooter (PG, SG, SF, PF, or C)
shooter_height	shooter height in inches
shot_clock	number of seconds remaining on shot clock at time shot was taken
shot_dist	shot distance from goal in feet
shot_number	number of shots taken previous to current shot
touch_time	number of seconds player has touched the ball prior to shot

Table 1. Potential Predictors for Modeling Shot Makes (* indicates derived variable)

DATA PREPARATIONS

Prior to analyzing the data, multiple steps were taken to prepare the data. The first step was filtering out implausible values (Table 2) and extreme outliers within the data (Table 3). In total 2,025 rows of data were removed (1.58% of original data). General data cleanup was then performed to ensure that the data was properly formatted for logistic regression. The primary benefit of data cleanup was the formatting of game clock values such that differences in field goal percentage (FG%) could be measured across time in the game.

REMOVAL OF IMPLAUSIBLE VALUES

Implausible Value	Description	# Observations Removed
Incorrectly labeled 3 point shot	pts_type = 3 & shot_dist < 22	872
Incorrectly labeled 2 point shot	pts_type = 2 & shot_dist > 23.75	539
Touch time prior to shot	touch_time < 0 or touch_time > 24	316
Player not in NBA	player in data set is not found on NBA Reference site	12

Table 2. Values Not Possible in an NBA Game

Due to physical constraints of NBA court design, certain shot distances are impossible for 2-point or 3-point shots. The 3-point design is an imperfect arc, with shorter distances at the corners of the court. A shot taken behind the 3-point line at the top of the arc must be at least 23.75 ft. from the goal, while a corner shot behind the 3-point line must be at least 22 ft. from the goal. In other words, all 3-point shots must be taken at least 22 ft. from the goal. Similarly, all 2-point shots must be taken less than 23.75 ft. from the goal. Setting these constraints likely omits most incorrectly labeled shots, but there is still the possibility that some shots falling between 22 ft. and 23.75 ft. are incorrectly labeled. Without exact floor location data these values can't be properly assessed.

REMOVAL OF OUTLYING SHOT TYPES

Outlier Shot Type	Description	# Observations Removed
Half Court Heave	shot_dist > 40	109
Breakaway Dunk	closest_def_dist > 20 & pts_type = 2	97
Wide Open Fast Break 3	closest_def_dist > 20 & pts_type = 3	77

Table 3. Shot Types Considered as Removable Outliers

Some shots included in the data set are not indicative of a typical NBA shot and occur only in rare situations. These shots have disproportionately higher or lower field goal percentages than the average NBA shot. In particular, shots from greater than 40 ft. have drastically lower FG% (2.6%) than a typical NBA shot. Meanwhile, breakaway dunks (99.1%) and wide open fast break threes (43.9%) are much more likely to be made than the typical 2-point (48.9%) or 3-point shot (35.3%). Removal of these values should improve model accuracy.

DATA CLEANING

The following data cleaning steps were taken prior to initial analysis of the data set:

- Fixed player names
  - concatenated first and last names for ease of reading + plotting
  - misspellings corrected (i.e. 'mnta ellis' → 'monta ellis')
- Separated original 'matchup' variable
  - Date component
  - Team names

- Location (Home/Away)
- Reformatted original 'game clock' variable
  - Time measure in seconds
  - New period second variable (0 indicates the end of a period)

EXPLORATORY DATA ANALYSIS

CLUTCH VS. NON-CLUTCH SHOTS

Before diving into prediction of whether a player will make a clutch shot, it is worth exploring how a clutch shot differs from a shot taken during a non-clutch situation. In general, players make shots at lower percentages as the game progresses (46.3% 1st quarter vs. 44.2% 4th quarter). This percentage is even lower during the final two minutes of close games (41.7%), but is not drastically different in the final minutes of games decided by greater than 5 points (44.4%). This may be an indication that clutch shots are an inherently different type of shot from their non-clutch counterparts. This idea will be explored further in the following figures. Figure 1, Figure 2, Figure 3, Figure 5 and Figure 7 all demonstrate differences and similarities between the distributions of predictor variables based upon clutch/non-clutch status. Figure 4, Figure 6, and Figure 8 all visualize how specific predictor variables affect FG% for both clutch and non-clutch situations.

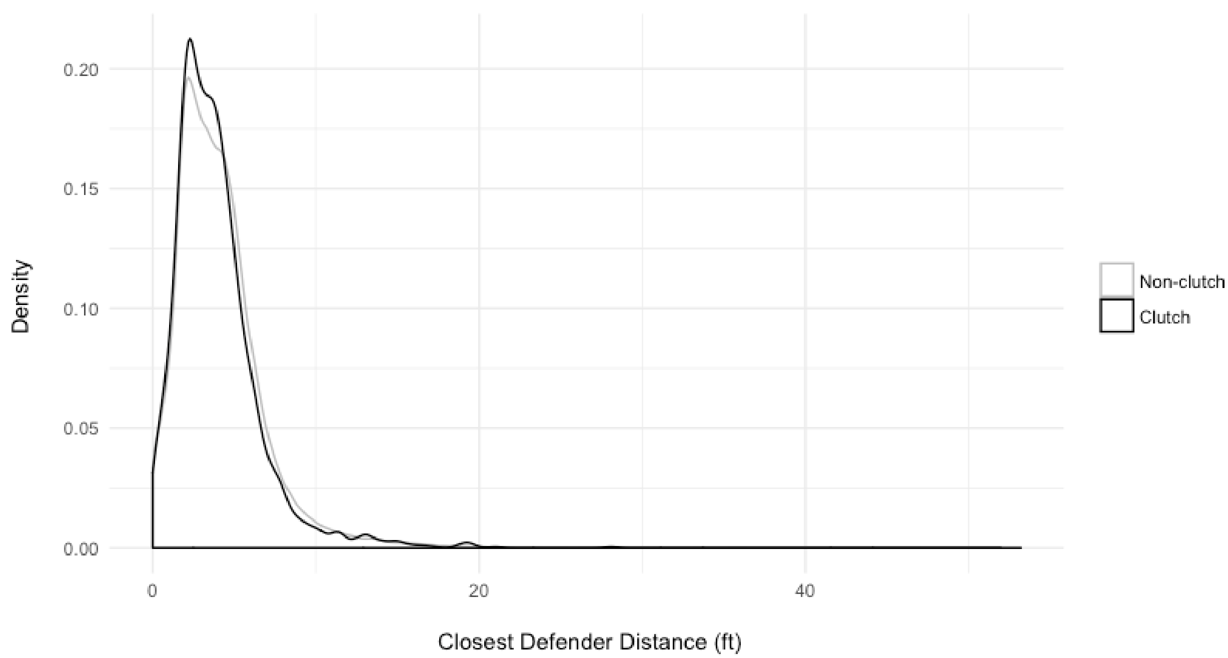

Figure 1. Kernel Density Estimates (KDE) of Closest Defender Distances

As we will see in the Model Building section of this paper, closest defender distance is a significant predictor of shot makes for both clutch and non-clutch shots. This is also the only significant predictor which does not drastically differ between clutch and non-clutch situations. The above distributions for both shot types are relatively similar.

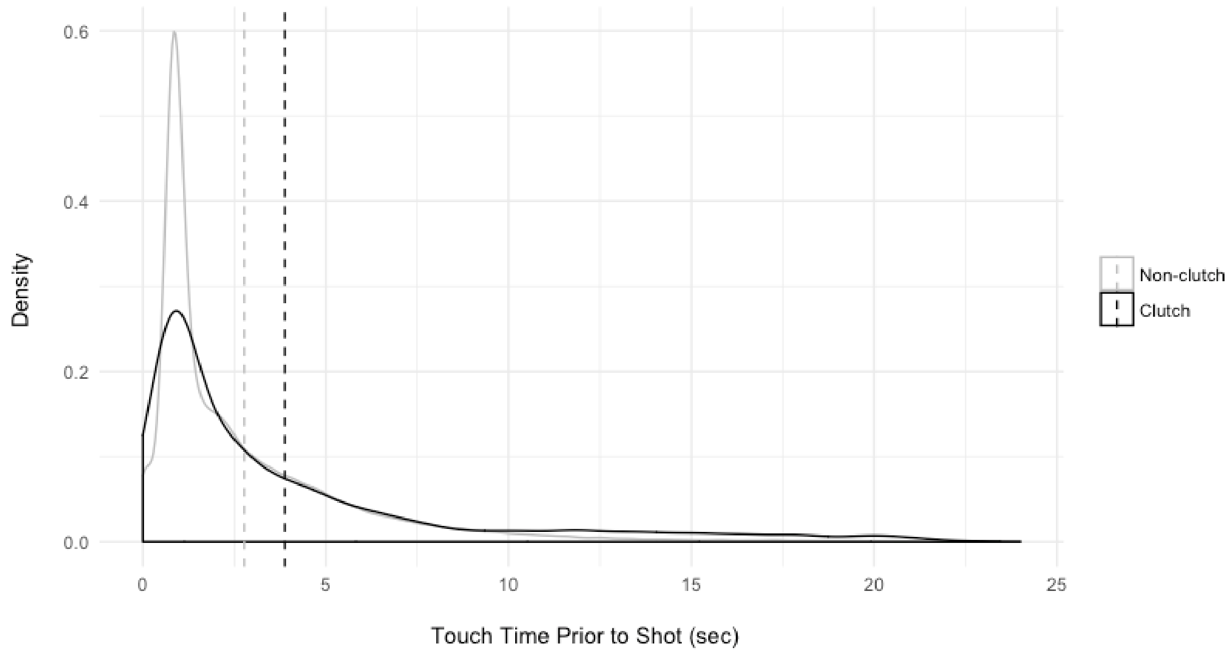

Figure 2. KDE of Shooter Touch Time

The above distributions indicate that players dribble out the clock or hold onto the ball longer during clutch situations than would be expected (3.89 seconds vs. 2.77 seconds for non-clutch). There are two primary coaching strategies that likely explain this. It may be primarily a result of players holding onto the ball to prevent the opposing team from getting another opportunity to score or a result of coaches advising players to take their time as the play develops. NBA teams typically design specific plays for the final minutes of games to take advantage of mismatched players or hot shooters. These plays are often more intricate than a typical NBA possession and take longer to unfold.

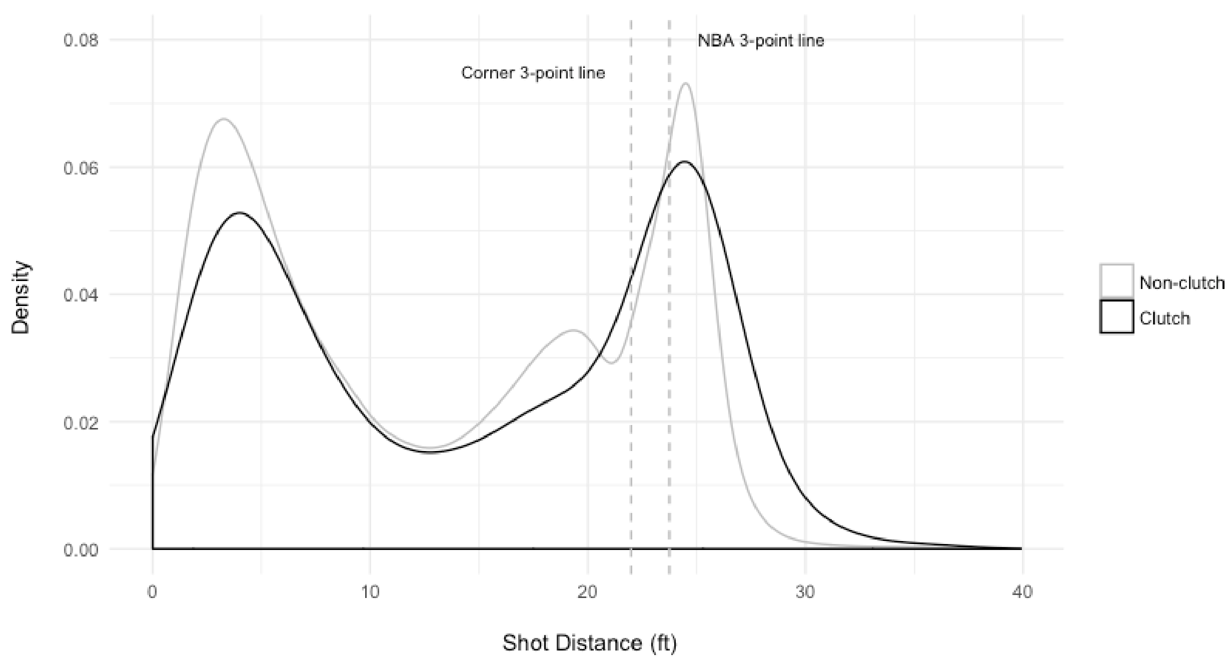

Figure 3. KDE of Shot Distance from Goal

Shots taken during crunch time are 3 pointers a disproportionate amount of the time. During non-clutch situations, a quarter of all shots are from 3-point distance as compared to over one third during clutch situations. Not only are shooters more likely to take 3-pointers during crunch time, they are also more likely to take these shots at longer distances. An average 3-pointer in non-clutch situations is from 24.6 ft., as opposed to 25.2 ft. for clutch situations. We see a strong bimodal distribution in both groups, in which players are significantly more likely to take shots near the basket or just past the 3-point line.

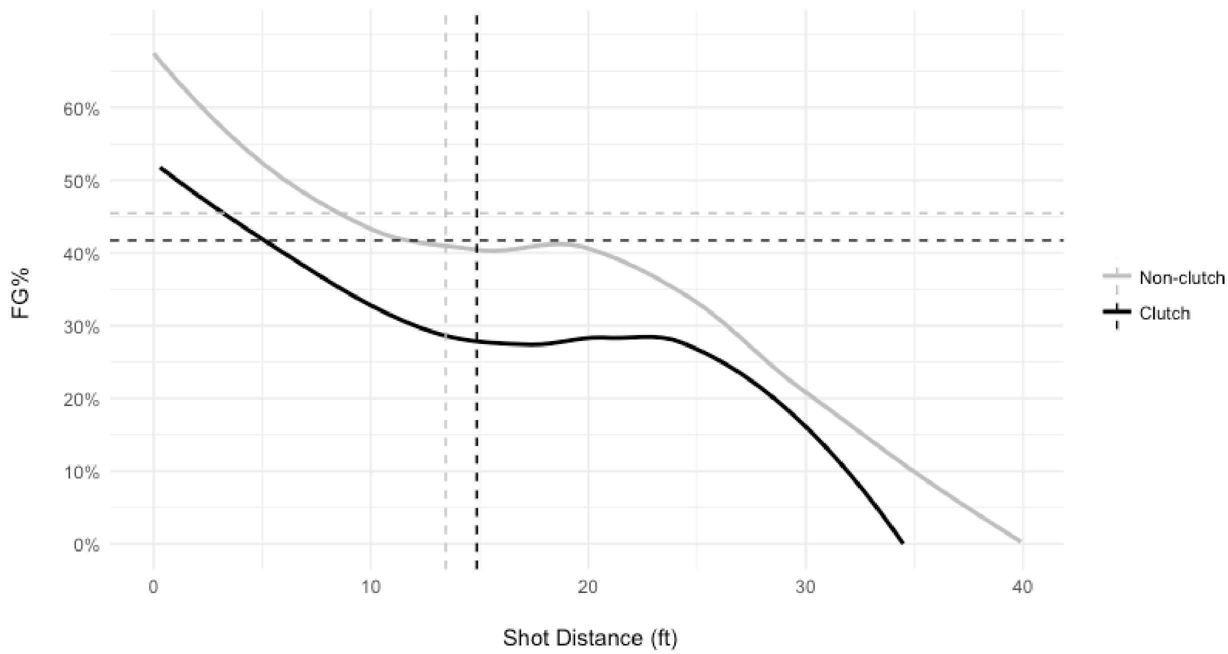

Figure 4. FG% Trend as Shot Distance Increases

Clutch shooters may be more likely to take 3-pointers from longer distances, but the affects of these longer distances don't appear to be stronger than would be expected. The trend in FG% is similar for both groups, with both following a sigmoidal curve. The flattening near 20 ft. is likely due to defenders guarding primarily at the rim of the goal or beyond the three-point line and not just inside the line (~20 ft.).

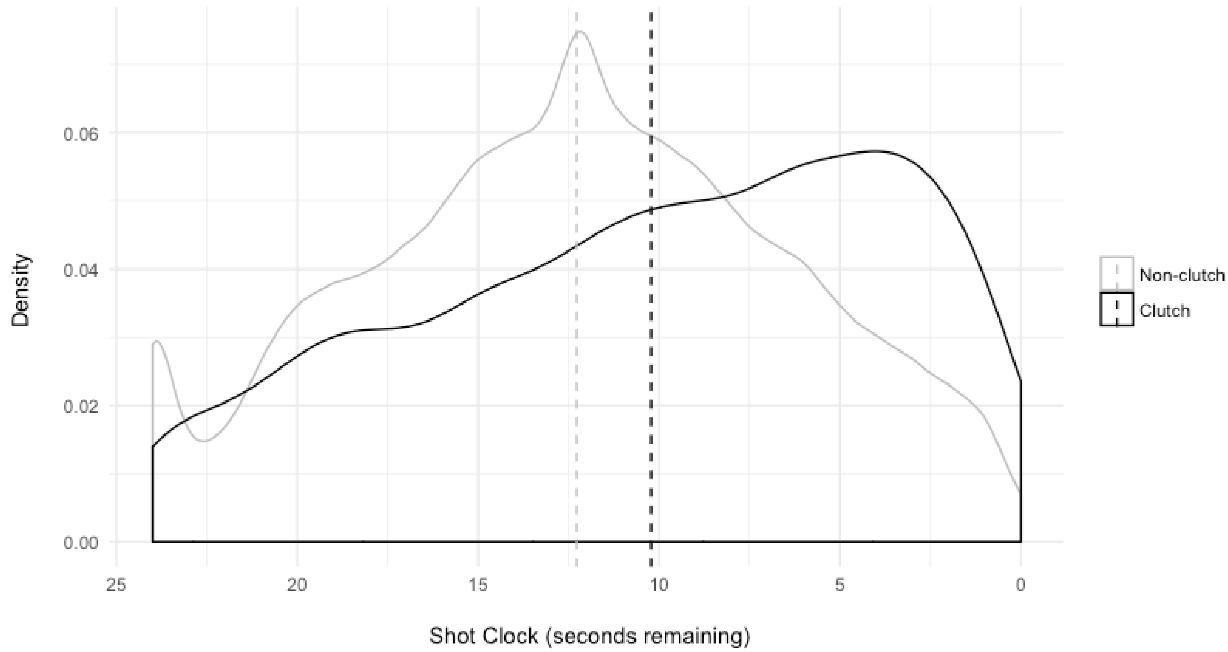

Figure 5. KDE of Time Remaining on Shot Clock at Time of Shot

The simplest explanation as to why clutch shots are made at significantly lower percentages is that high pressure shots at the end of games are much more likely to be taken late in the shot clock. This may be related to the explanation offered for Figure 2, in which coaching strategy strongly affects when the shot is taken.

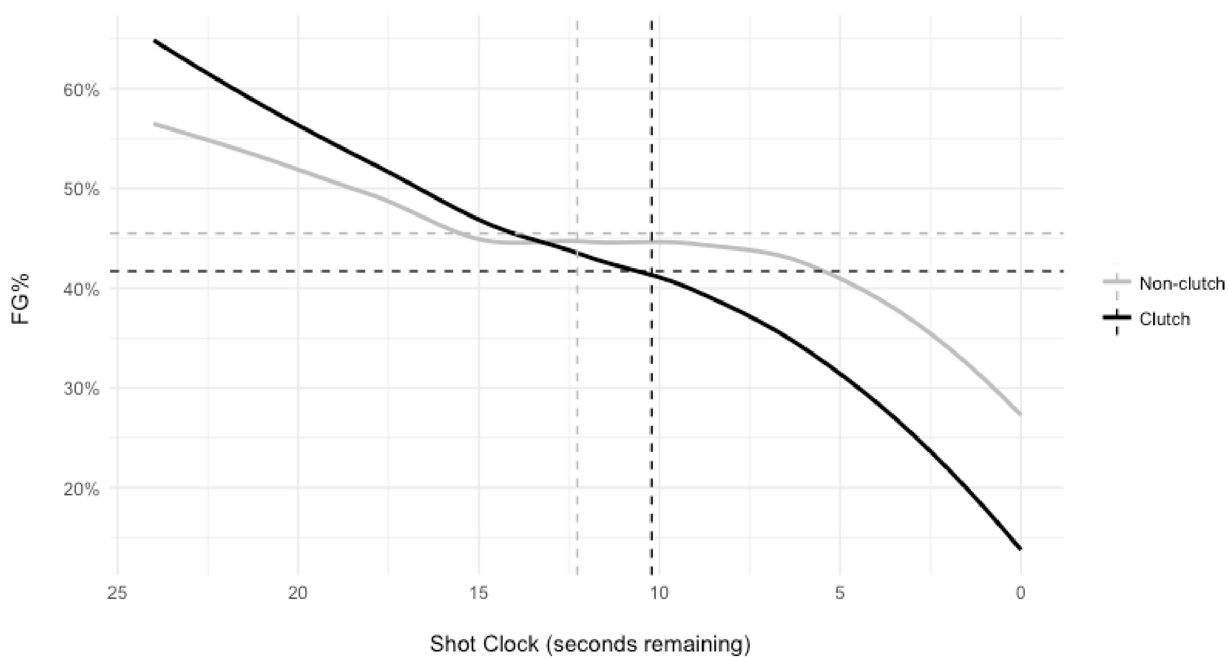

Figure 6. FG% Trend as the Shot Clock Winds Down

Shots taken at the beginning of the shot clock are likely wide open shots resulting from fast breaks or set out of bounds plays, while shots at the end of the shot clock are likely desperation shots in which no players were previously open or the opposing team played particularly good defense.

Interestingly, for clutch shooters shooting performance decreases steadily as the shot clock time decreases, while non-clutch performance levels off somewhat in the middle of the shot clock. I believe this may be an artifact of the last shot of the game. Shots taken in the last seconds of a game may occur early in the shot clock, but are still less likely to be made. A team may regain possession of the ball with 3 seconds remaining in the game and the shot clock would reset to 24 seconds. In other words, at the end of the game there are two possible time constraints, shot clock and game clock, that may not be influential at the same time.

PLAYER FATIGUE

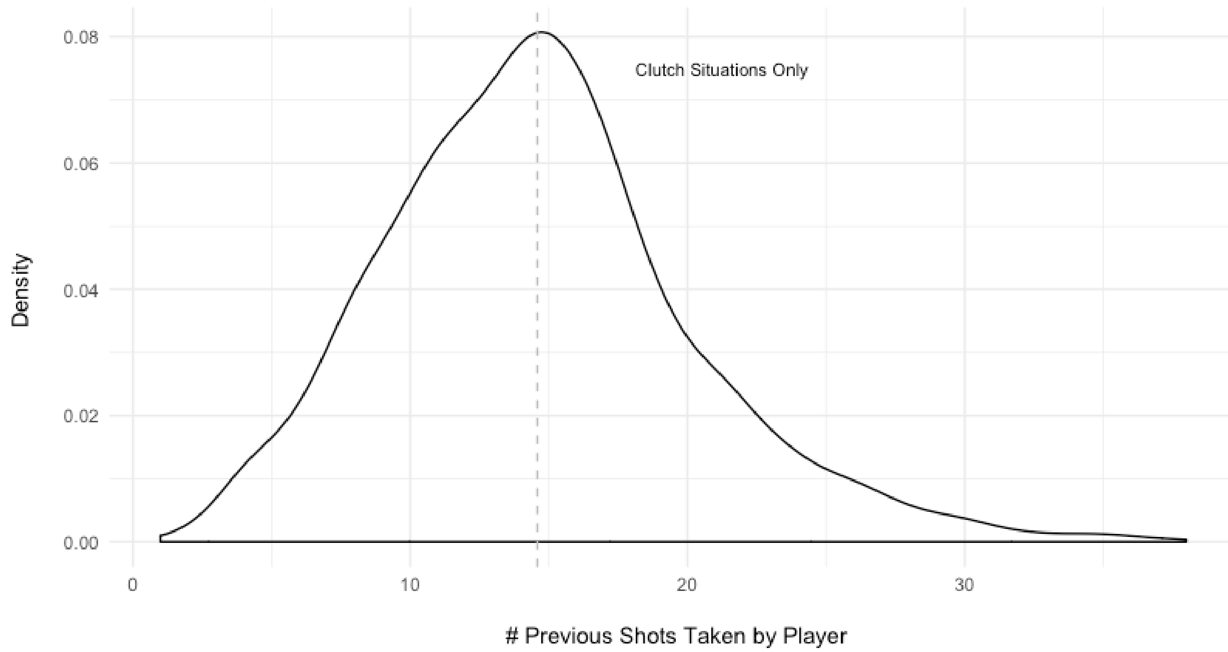

Figure 7. KDE of Player Shot Number within Game

It makes intuitive sense for a team to have their top shooters take shots in clutch situations, but it is still interesting to note just how many shots a player is likely to have taken previously. On average, a player will have taken 13.6 shots previous to taking a clutch shot. It is very unlikely for a team to have a player who is ‘cold’ take a shot during late game situations, with only 5.7% of clutch shots being taken by a player with 5 or less previous shots.

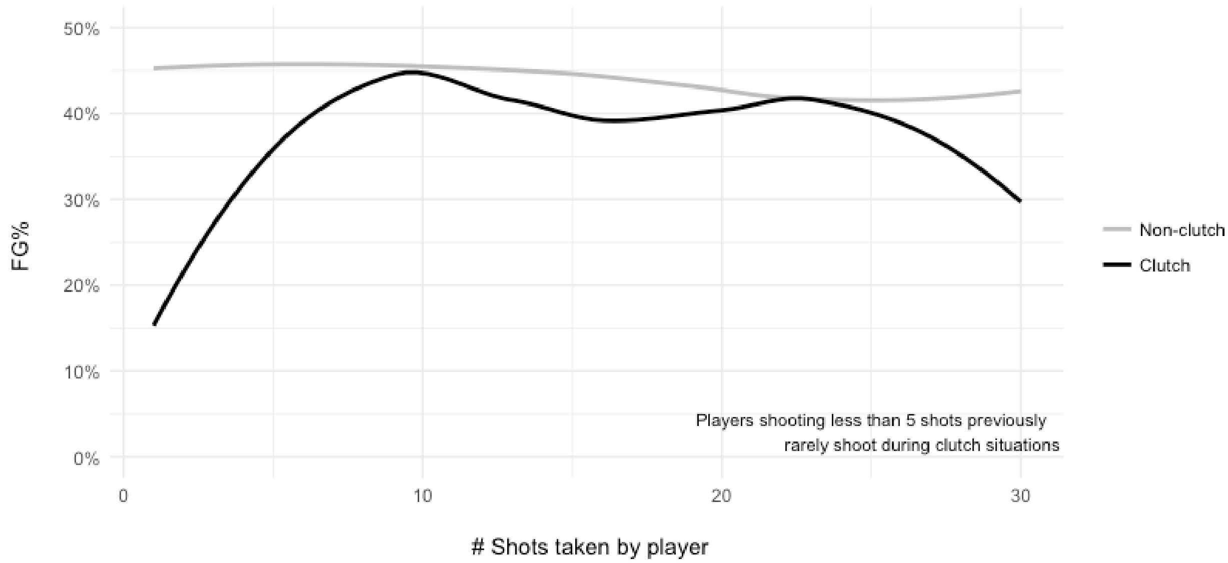

Figure 8. FG% Trend as the Number of Shots Taken by a Player Increases

Using number of shots taken as a proxy for player fatigue appears to be a mediocre metric of shot success. There isn't a general trend in FG% as the number of shots taken previously increases and there is too much uncertainty at the tail ends of the clutch distribution to have a fair assessment. A more direct approach to measuring fatigue (i.e. number of minutes played at time of shot) would be a better metric to include in future models.

CLUTCH SHOOTERS

Figure 9 below provides an assessment of how high volume clutch shooters perform relative to the average. In order to be considered 'high volume' a player had to have taken at least 10 clutch shots during the period of the season included in the NBA Shots data set. Size of points indicates the volume of clutch shots taken.

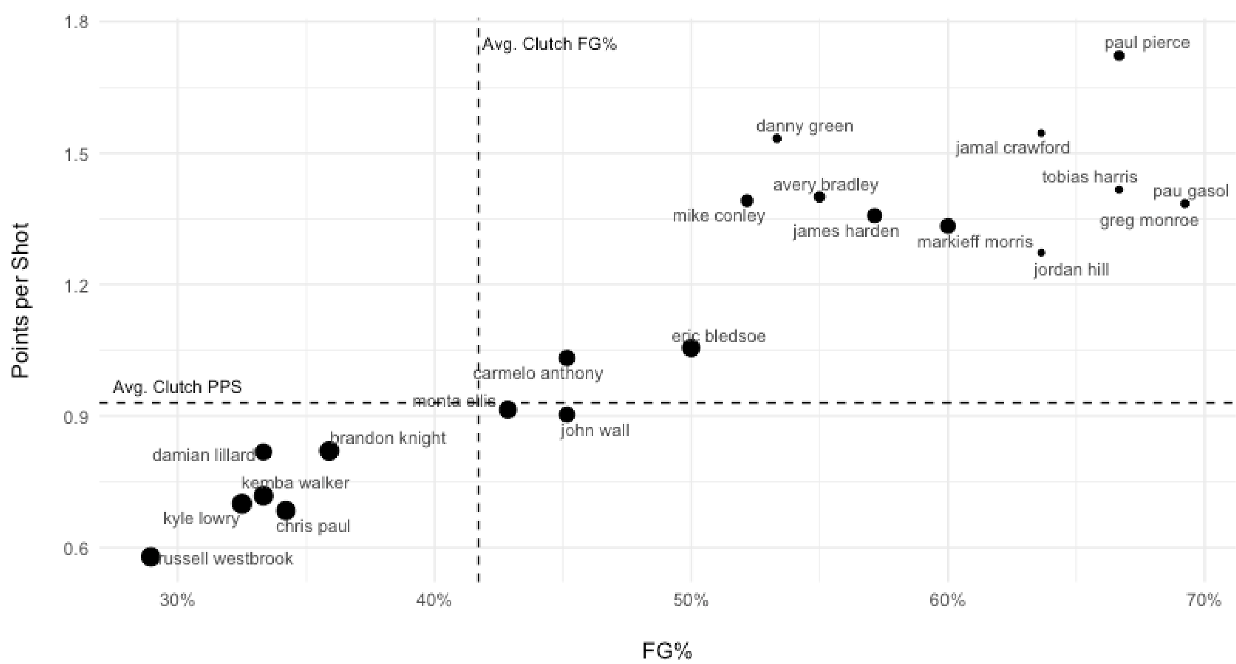

Figure 9. Performance of High Volume Clutch Shooters

Many of the highest volume clutch shooters are situated in the bottom left corner of Figure 9, below the average in both FG% and points per shot. In contrast, players who perform well above expectation are more likely to have smaller dots indicating lower volume of clutch shooting. This may be a case of regression towards the mean as players take more shots during clutch situations. Kemba Walker and Damian Lillard are both generally consider clutch game-winning shooters, but both are below average in the above figure. The limited number of clutch shots included in this data set (n = 2,096) makes assessment at the player level difficult. Including clutch shooting across multiple seasons would give a better picture of which players truly are great during crunch time.

FEATURE ENGINEERING

In order to more accurately model shot makes, two derived attributes were created from other variables within the data set. While the inclusion of these two predictors increased the accuracy of the final models, they also served to reduce issues of multicolinearity seen in Table 4.

MULTICOLINEARITY OF PREDICTORS

There are three primary positive covariances of concern. The correlation between shooter and defender heights is easily avoided by deriving a height differential metric (Figure 10) that measures how much taller a shooter or defender is from his opponent. The relationship between number of dribbles prior to the shot and touch time is strongly positive and immediately noticeable in Table 4 below. This is entirely reasonable, as the values are essentially both measuring the amount of time a player has the ball before he shoots. Instead of measuring the amount of time a player has the ball prior to the shot, a catch and shoot variable was derived (Figure 11) that indicates whether the player dribbled prior to shooting. This variable has greater predictive power than either of dribbles or touch time and solves the issue of colinearity between the two. The relationship between shot distance and closest defender distance is more complicated. Previous analyses (Cen, et al., 2015, Zhang 2016) have suggested an 'openness' metric that encompasses both distance to the goal and distance from defenders, but without exact floor locations of where the shot was taken this is not possible. I have opted to retain both of these variables, as they are both significant predictors of shot outcome.

The negative correlations between shooter height and dribbles/touch time is also problematic, but is reduced after shooter height is removed in favor of height differential and dribbles and touch time are omitted in favor of the catch and shoot variable.

	Field Goal Made?	Shot Number	Shot Clock	Dribbles	Touch Time	Shot Distance	Closest Defender Dist.	Shooter Height	Defender Height
Field Goal Made?	1								
Shot Number	-0.01	1							
Shot Clock	0.10	-0.04	1						
Dribbles	-0.04	0.14	-0.10	1					
Touch Time	-0.05	0.15	-0.15	0.93	1				
Shot Distance	-0.19	0.01	-0.18	-0.08	-0.08	1			
Closest Defender Dist.	0.00	-0.04	0.02	-0.15	-0.17	0.53	1		
Shooter Height	0.05	-0.07	0.00	-0.39	-0.32	-0.21	-0.05	1	
Defender Height	0.02	-0.02	-0.01	-0.10	-0.07	-0.24	-0.04	0.39	1

Table 4. Correlations between numeric variables in NBA Shots data set

HEIGHT DIFFERENTIAL

Figure 10 below demonstrates that height differential increases during clutch situations.

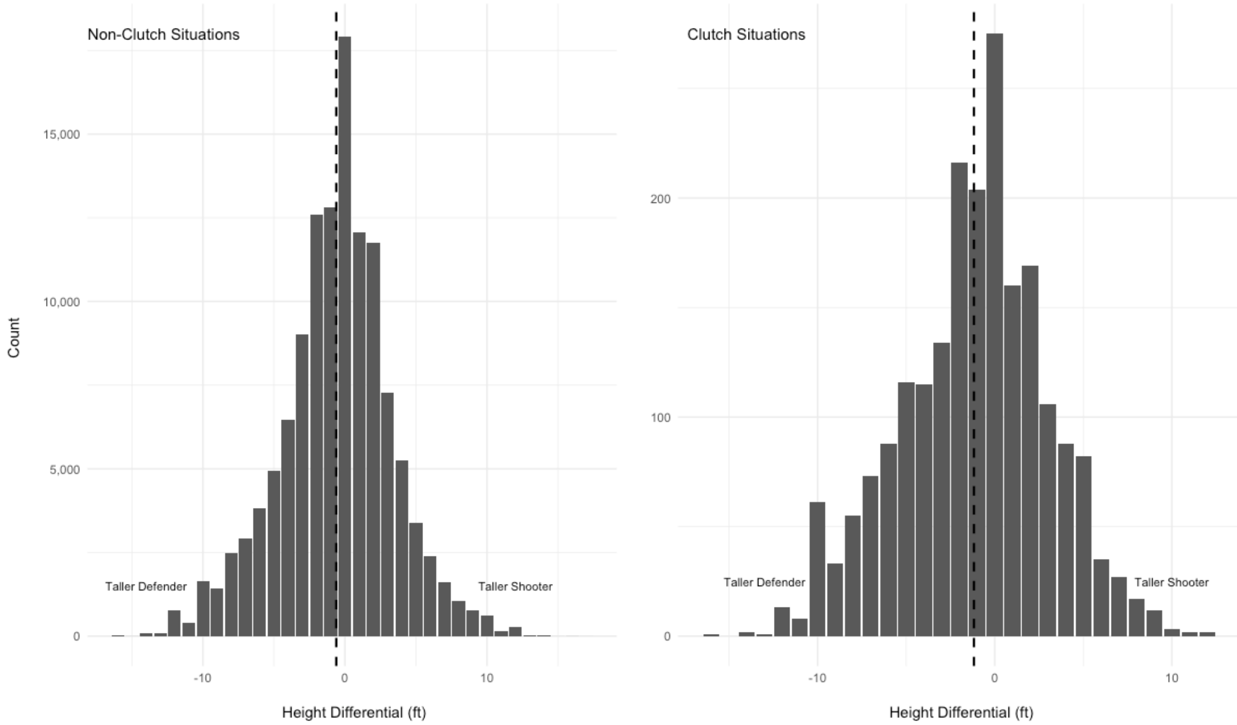

Figure 10. Height Differential Distributions of Non-Clutch and Clutch Shots

During clutch situations, the difference in height between defender and shooter is greater (1.18 inches on average) than during non-clutch situations (0.61 inches on average). While both distributions in Figure 10 are approximately normal, there is a slight left skew in the clutch shooting distribution. As the clock nears the end of regulation, opposing teams are more likely to swap out taller defenders to guard star players. Teams likely call timeouts or make in game adjustments to have taller, better defenders on the court.

CATCH AND SHOOT

Figure 11 below shows the increases in average FG% that occur when a player does not dribble prior to taking a shot.

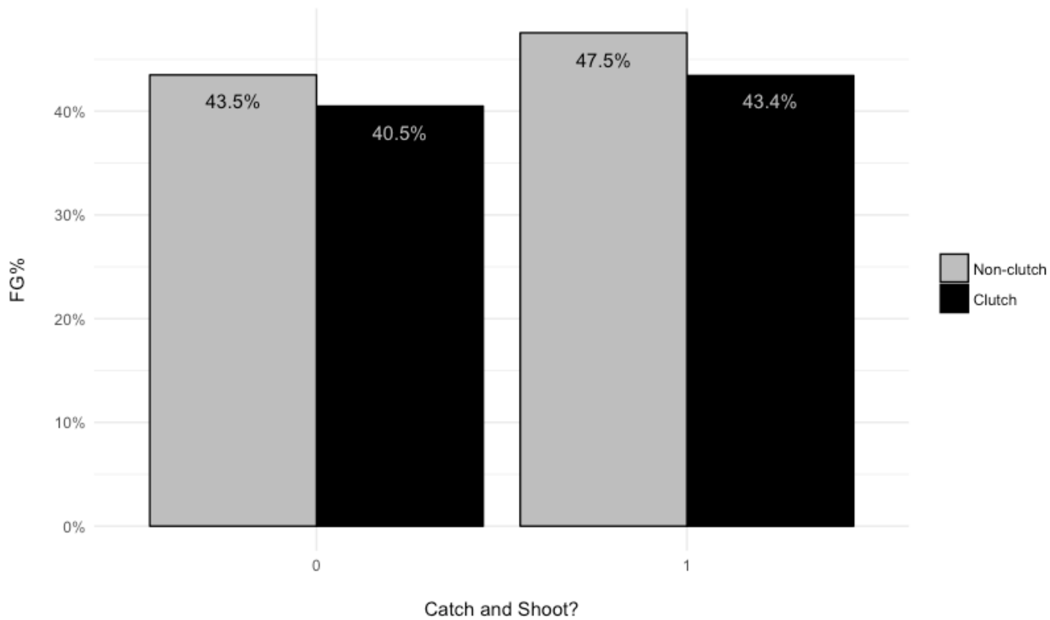

Figure 11. FG% by Catch and Shoot Scenario

We can clearly see that for both non-clutch and clutch situations shooting immediately upon catching the ball is a higher percentage shot than taking dribbles prior to shooting. Not only is it more difficult for a player to accurately align his shot trajectory while dribbling, players are likely more open in catch and shoot situations, either by coming around a screen for a shot or being open due to the passing of the ball around the court.

MODEL BUILDING

The following SAS® code uses the SURVEYSELECT procedure to split data into training and validation data sets for model building and assessment. This method uses stratified random sampling specified in the METHOD option of PROC SURVEYSELECT to ensure that the outcome variable is evenly distributed between training and test sets. PROC SURVEYSELECT was run on both the non-clutch and clutch data sets, but only the non-clutch example is included for brevity:

```
proc surveyselect data=nba.nba_shots out=nba.nba_split
   method=srs samprate=0.7 outall noprint;
run;

data nba.nba_shots_train;
   set nba.nba_split;
   if selected=1;
   drop selected;
run;

data nba.nba_shots_test;
   set nba.nba_split;
   if selected=0;
   drop selected;
run;
```

The following code was run using the LOGISTIC procedure to create models for the training data and scoring data sets for the test data. PROC LOGISTIC was run for both the full data and shots taken only during clutch situations, with only the non-clutch example appearing below. In both cases the predictors included in the MODEL statement were the same. In order to determine the optimal model, forward stepwise selection was performed using the SELECTION option of the MODEL statement. The CLASS statement creates effect coding design matrices for the listed classification variables. The SCORE statement indicates that the test data should be used for scoring purposes:

```
proc logistic data=nba.nba_shots_train;
   class location pos;
   model fgm (event = '1') = shot_clock touch_time location shot_number
      shot_dist close_def_dist height_diff pos catch_shoot/ctable pprob=0.5
      selection=forward rsquare link=logit expb stb;
   score data=nba.nba_shots_test out=nba.full_fit;
   oddsratio shot_clock;
   oddsratio location;
   oddsratio shot_number;
   oddsratio shot_dist;
   oddsratio close_def_dist;
   oddsratio height_diff;
   oddsratio pos;
   oddsratio catch_shoot;
run;
```

FORWARD STEPWISE SELECTION

Table 5 and Table 6 below contain the variables selected as predictors after forward stepwise selection for the non-clutch and clutch model, respectively.

Summary of Forward Selection					
Step	Effect Entered	DF	Number In	Score Chi-Square	Pr > ChiSq
1	shot_dist	1	1	4555.8635	<.0001
2	close_def_dist	1	2	1903.4292	<.0001
3	catch_shoot	1	3	421.6331	<.0001
4	shot_clock	1	4	282.0314	<.0001
5	height_diff	1	5	82.1678	<.0001
6	pos	4	6	148.5303	<.0001
7	touch_time	1	7	36.5447	<.0001
8	shot_number	1	8	6.4001	0.0114
9	location	1	9	5.4108	0.0200

Table 5. Non-Clutch Model Forward Stepwise Selection

The stepwise selection method for the non-clutch model selected all of the predictors considered in the PROC LOGISTIC MODEL statement. Each of the 9 variables above were statistically significant predictors of shot success ($p < 0.05$). While logistic regression does not have an exact variable importance metric, comparison of standardized coefficients indicated that distance of the shot from the goal has the most effect upon shot success.

Summary of Forward Selection					
Step	Effect Entered	DF	Number In	Score Chi-Square	Pr > ChiSq
1	shot_dist	1	1	97.8687	<.0001
2	close_def_dist	1	2	29.7098	<.0001
3	shot_clock	1	3	24.7654	<.0001
4	height_diff	1	4	6.2582	0.0124

Table 6. Clutch Model Forward Stepwise Selection

The stepwise selection method for clutch only shots did not include all of the variables seen in the non-clutch model. Only 4 of the 9 potential predictors were added to the model based upon (Pr > ChiSq) < 0.05. This is probably due to the smaller size of the clutch shooting data and not an indication that the previously significant predictors (catch_shoot, pos, etc.) have no predictive power in modeling clutch shot outcomes.

ODDS RATIOS

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
shot_clock	1.016	1.013	1.018
touch_time	0.984	0.979	0.989
location A vs H	0.973	0.951	0.996
shot_number	1.003	1.001	1.006
shot_dist	0.936	0.934	0.937
close_def_dist	1.113	1.107	1.120
height_diff	1.022	1.019	1.026
pos C vs SG	0.812	0.779	0.847
pos PF vs SG	0.860	0.830	0.892
pos PG vs SG	1.047	1.011	1.084
pos SF vs SG	0.893	0.860	0.927
catch_shoot	1.215	1.179	1.252

Table 7. Non-Clutch Model Odds Ratios

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
shot_clock	1.045	1.027	1.063
shot_dist	0.932	0.919	0.945
close_def_dist	1.135	1.079	1.193
height_diff	1.034	1.007	1.061

Table 8. Clutch Model Odds Ratios

The increase in odds for each of the common predictors (shot_dist, close_def_dist, shot_clock) between the two models are relatively similar. In other words, for each unit increase in these predictors, we would expect similar increases in the odds of shot success for both non-clutch and clutch models. This may be a good indicator that modeling of non-clutch and clutch shots do not differ as much as may be expected.

MODEL ASSESSMENTS

PERFORMANCE METRICS

Table 9. Non-Clutch Model PerformanceTable 9 and Table 10 below give a full assessment of both models and compare performance between training and test data.

	Training Data	Test Data
Accuracy	0.61	0.61
Sensitivity	0.46	0.46
Specificity	0.74	0.74
Positive Predictive Value	0.59	0.6
Negative Predictive Value	0.62	0.62

Table 9. Non-Clutch Model Performance

	Training Data	Test Data
Accuracy	0.65	0.66
Sensitivity	0.42	0.45
Specificity	0.82	0.82
Positive Predictive Value	0.62	0.65
Negative Predictive Value	0.67	0.67

Table 10. Clutch Model Performance

While prediction accuracies of 61% (non-clutch shooting) and 66% (clutch shooting) are not necessarily impressive, we need to keep in mind what a null model would predict. Non-clutch shots are made 45.4% of the time, so a ‘dumb’ model that predicts only misses would be correct 54.6% of the time. For clutch shots, 41.7% of all shots are made, so a ‘dumb’ model would be correct 59.3% of the time.

We should also take a look at how the model is missing outcomes. For both the non-clutch and clutch shots, the models do a poor job of predicting shots that were actually makes (sensitivities of 0.46 and 0.45, respectively). The models do a relatively good job of predicting shots that were actually missed (specificities of 0.74 and 0.82, respectively).

DOUBLE DENSITY PLOTS

Figure 12 shows how scoring values in the model test sets compare to the actual distributions of the binary outcomes (missed shot/made shot). For both models I have assigned the cutoff probability value to be 0.5, which means that all prediction values at or above 0.5 are classified as positive cases (made shot) by the model.

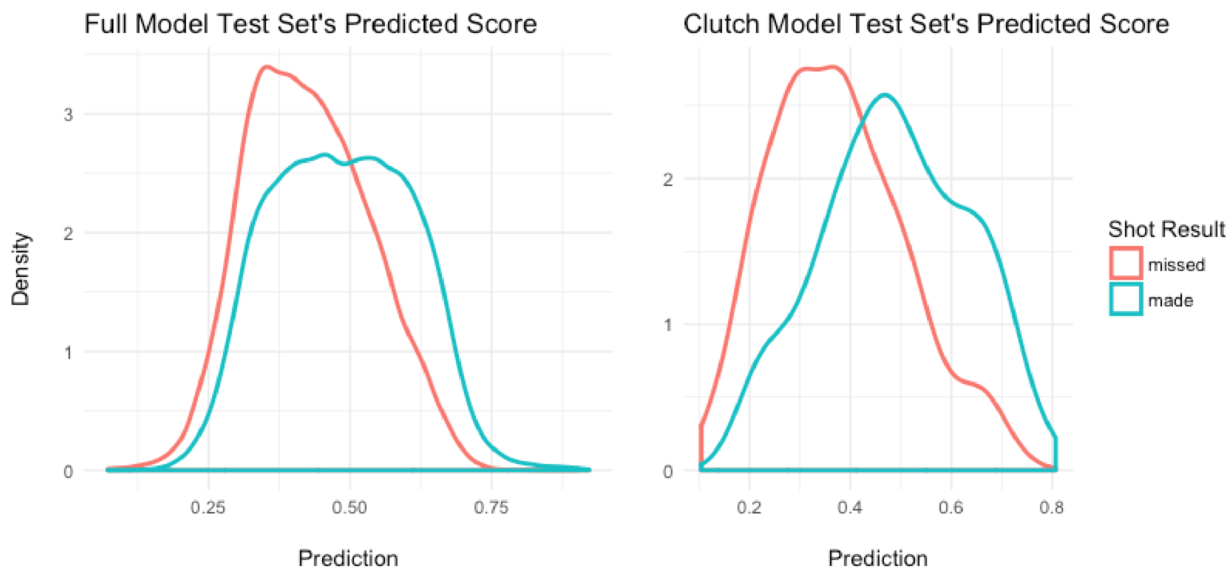

Figure 12. Double Density Plots for Non-Clutch and Clutch Model Predictions

Ideally, a double density plot for the distributions of binary outcomes in a classification model should be well separated across prediction values. Plots that have largely separated distributions indicate that the model scoring does an adequate job of separating positive cases from negative cases, while similar distributions indicate that the model has difficulty in separating positive and negative cases.

In Figure 12, we see limited separation when modeling non-clutch shots. The model has trouble distinguishing between missed and made shots that are not at the tail ends of the prediction score distribution.

The clutch shooting model did a much better job of predicting positive and negative cases, as seen in the greater separation between their distributions. This is expected, as Table 10. Clutch Model Performance shows higher accuracy for the clutch model vs. the non-clutch model.

CONCLUSION

While both the non-clutch and clutch models improved upon a null model in which all shots are predicted as misses, there is room for improvement.

Ideas for how modeling clutch shooting could be improved:

- More Data

  The NBA Shot data set contained ~2k shots that met my definition of clutch. Additional data from more recent seasons would give a better overall picture of clutch shooting. While I doubt significant predictors would vary with additional information, the general trends in clutch shooting would be clearer and aggregation at the team and player levels would be more accurate.

- New Features

  As discussed in previous sections of the paper, new features that explain variation in shot outcomes could be included in the future. Specifically, a predictor that explores how ‘open’ a player is in relation to both distance from the goal and the nearest defender would be beneficial. Shot and closest defender distances were both significant predictors for the clutch model, but in a sense they both partially explain how ‘open’ a shot is. Derived features from x,y coordinates of where a shot was taken on the floor would be beneficial.

- Separate Models

  2-point and 3-point shots are inherently different shots with different predictor variable distributions. Outside of the obvious shot distance differences between 2 and 3 pointers, these shot types are defended differently and often occur at differing times in the game and shot clock.

  Being limited to only one partial season of data, I did not model based upon shot points type (2 vs. 3 pointer), as this would have severely restricted the number of clutch shots for each model. I suspect that the accuracy of each of these separate models would be higher than that of the current models.

REFERENCES

Cen, R., Chase H., Pena-Lobel, C. and Silberwasser, D. NBA Shot Prediction and Analysis. 2015. Accessed July 25, 2017. https://hwchase17.github.io/sportvu.

Kaggle. NBA Shot Logs. 2016. Accessed April 20, 2017. https://www.kaggle.com/dansbecker/nba-shot-logs.

Kaggle. Adding Personal Players Information. 2016. Accessed April 20, 2017. https://www.kaggle.com/dansbecker/nba-shot-logs/discussion/25806.

Zhang, Gilbert. Predicting NBA Shot Results. 2016. Accessed July 30, 2017. https://github.com/gyzhang328/Predicting-NBA-Shot-Results/blob/master/Presentation.pdf.

ACKNOWLEDGMENTS

I would like to thank Drs. Goutam Chakraborty and Miriam McGaugh of Oklahoma State University for their guidance and insight in the writing of this paper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Logan Edmonds
Oklahoma State University
918-759-7648
loganpe@okstate.edu