Exploring Expected Points and its Applications in the NBA

Lucca Ferraz

April 30, 2024

1 Introduction

How can we use the concept of shot quality in basketball to best measure team and player performance in the NBA? One of the biggest advancements in modern-day soccer analytics has been the proliferation of expected goals metrics. For those unfamiliar with the concept, soccer analysts realized that oftentimes the final score can be incredibly misleading. In such a low-scoring game, variance tends to run wild, as unexpected results where one team dominates the entire game and loses are commonplace. To combat the noisy stat that is goals, analysts came up with the idea of expected goals. Using multiple factors of the shot, such as distance, angle, shot type, defenders, et cetera, they could predict how likely a given shot was to go in. Once they have this number, they could look at what players or teams created the best shots, or which players / teams were the most efficient in finishing their chances. The concept has rapidly spread throughout the sport analytics world, with the stat being adopted in hockey as well. For the purposes of this project, I wanted to take a similar outlook to modeling shot quality in basketball.

Synergy Sports has their own Shot Quality metric, but they keep the formula tightly guarded and as such it is difficult for members of the public to interpret. Pretty much all advanced measures of shot quality like the measure Synergy has are locked behind paywalls, making it difficult for the average NBA fan to gain access.

The idea can also be incredibly valuable to NBA front offices when it comes to team-building and roster development. Knowing the underlying value of shots, rather than simply judging them based on the binary outcome of miss-or-make, can provide a much more holistic view of a player's true impact on the court. For example, an incredibly skilled passer on a team with poor shooters may have low raw assist numbers, but he is likely still creating incredibly high-value shots for his team even though they can't convert the looks. Shot quality can also be important in measuring scheme and lineup success, especially with small sample sizes. Perhaps a certain 5-man starting lineup has performed badly in their first few games playing together, but if they are generating quality shots and stopping their opponent from doing the same eventually the shooting "luck" should balance out.

2 Data

For this project, I utilized NBA play-by-play data for the 2023-24 season from the HoopR package of the SportsDataVerse. I also utilized player and team box score data for the same season, all through the HoopR package.

3 Methods

3.1 Shot Quality Metric

When tackling this question myself, the main factors I wanted to consider were shot type and distance from the rim. Oftentimes I have seen public work in this domain include quarter or time remaining in the game, but I felt these do not tangibly impact a given shot any more than simply using the type of shot and distance. To construct my model, I sorted the shot types into five groups: dunks, layups, hook shots, free throws, and jump shots. For distance metrics, I calculated the horizontal distance from the rim, the vertical distance from the rim, and the total distance from the rim, so the model had

Coefficient	Value
Intercept	2.115
Free Throw	-0.43
Hook Shot	-2.06
Jump Shot	-1.93
Layup	-1.75
Other Shot	-2.19
X Distance	-0.004
Y Distance	-0.006
Total Distance	-0.03

Table 1: Logistic Regression Summary

a holistic view of where the shot was taken. I then implemented a logistic regression to predict the probability of a given shot going in. Once I had the probability, I could multiply it by the value of the shot (3 for three-pointers, 2 for twos, 1 for free throws) to get the Expected Point value of any shot. Knowing the Expected Point value of a shot and the result of the shot, I simply subtracted Actual Points minus Expected Points to get the stat Points Over Expectation. The output of the logistic regression model is included in Table 1 above.

3.2 Regularized Adjusted Plus-Minus Model

After producing the expected points stat, I wanted to take it a step further in terms of its application to player evaluation in particular. Beyond just grouping players based on their expected points (which would mostly just measure who takes the most shots) or their points over expectation (a good metric that will be covered more later, but only has applications to scoring), I wanted to measure individual player impact on team expected points. In order to do this, I implemented a regularized ridge regression model. This involved taking the 10 players on a court over given periods of time (stints), finding the expected point differential (total expected points for team A minus total expected points for team B), and dividing by minutes played so we have a standardized measure of expected score differential per minute for each stint. Once I had that, I implemented a ridge regression to find the players with the highest coefficients (utilizing the lambda that minimized the MSE), meaning that they most positively impacted their team's shot quality on both sides of the ball when in the game.

3.3 Simulating NBA Playoffs Using Points Over Expectation

With playoff season in the NBA in full swing, I wanted to utilize my Expected Points metric to help inform predictions for how the bracket will play out. As will be further discussed in the Results section, the correlation between Expected Point differential and actual wins was not great, whereas the correlation between Points Over Expectation and actual wins was much better. To predict game outcomes using POE differential, I utilized the Bernoulli Win-Loss outcome model

$$Z_i \sim Bernoulli(e^{\eta_i}/(1+e^{\eta_i}),$$

$$\text{where } \eta_i = PointsOverExpectationPerGame/10.$$

To compare two teams, I subtracted Team B's POE Differential from Team A's POE Differential to get my η_i to plug into the equation above. Once I had the probability for a given team winning a head-to-head matchup in one game (through Z_i), I could simulate an NBA playoff series repeating until one team had 4 wins. For each matchup, I ran the simulation 10000 times, using a matrix to store the results (4-0 Team A, 4-3 Team B, etc). To fill out the bracket, I simply took the most common outcome from each simulation.

4 Results

4.1 Top Teams

4.1.1 Expected Points

Below is a ranking showing the top teams for 2023-24 in terms of Expected Points Scored. The top 5 teams were,

1. Indiana Pacers (8839)
2. Atlanta Hawks (8771)
3. Utah Jazz (8743)
4. Los Angeles Lakers (8691)
5. San Antonio Spurs (8567)

We can see through these results that more Expected Points doesn't necessarily lead to a better offense, as while the Pacers, Hawks, and Lakers all boast top 6 offenses in Points Per Game, the Jazz are much more middle-of-the-pack and the Spurs are a bottom 10 offense this season.

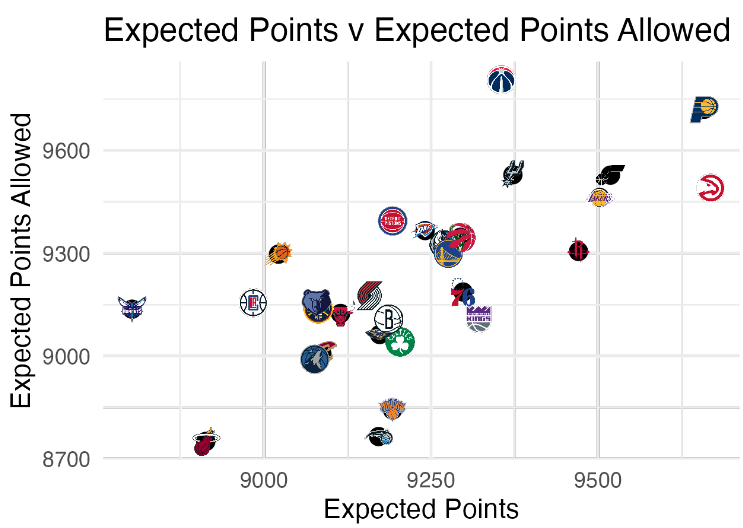

Figure 1: Expected Points vs Expected Points Allowed

The plot above shows every NBA team's expected points compared against their expected points allowed. You would expect the best teams to be in the lower right quadrant, having more expected points and less expected points allowed. However, there doesn't appear to be any pattern as to where the best teams in the league fall, as the Celtics and Thunder are both middle-of-the-pack and the Timberwolves struggling in relation to expected points.

4.1.2 Points over Expectation

After seeing that basic Expected Points didn't seem to relate to team strength that well, I decided to look further into Points Over Expectation in relation to measuring team strength. Below are the top 5 teams in terms of Points Over Expectation:

1. Boston Celtics (611)
2. Oklahoma City Thunder (592)
3. Los Angeles Clippers (515)
4. Phoenix Suns (480)

5. Indiana Pacers (473)

This gives more reasonable results when ranking the top offenses in the NBA today. We can see that the cream of the crop tends to outperform their Expected Points by wide margins, likely due to superior talent and shot-making ability.

4.2 Predictive Power?

4.2.1 Correlation to Winning

Statistic	Correlation
First-Half Win Percent	0.803
First-Half Point Differential	0.799
First-Half Points Over Expectation Differential	0.756
First-Half Expected Point Differential	0.431

Table 2: Comparison of Statistics

In order to determine which (if any) of the Expected Point statistics were good enough predictors to attempt to build a playoff bracket with, I separated the NBA season into 2 equal-sized chunks of 41 games each. I found each teams second-half win percentage, and compared the correlations for each predictor statistic in question to see which ones did the best job in predicting the out-of-sample second-half of the season. Table 2 above contains each possible predictor and its correlation coefficient with 2nd-half win percentage. We can see that while basic stats like win percent and point differential are still the best predictors for second-half win percentage, POE (Points Over Expectation) differential remains not too far behind and is a far better predictor than Expected Point Differential. With that in mind, I continued making my playoff predictions with POE differential albeit with a grain of salt.

4.2.2 Model Predictions

Figure 2 below contains the playoff predictions based on POE Differential and the simulation described in 3.3. While most of the bracket seems reasonable, we can already see the pitfalls of using POE Differential as a predictor, as the model predicted the Phoenix Suns to reach the Conference Finals and they have already been swept in the first round of the playoffs.

Figure 2: Playoff Predictions Based on POE diff

4.3 Top Players

4.3.1 Points Over Expectation

After finding that Points Over Expectation carried a reasonably strong correlation with winning at the team level, I wanted to rank individual players based on their POE to see who the best scorers / shot-makers in the game today are. Below is a list of the top 5 players in the NBA this year based on POE:

1. Luka Doncic (385)
2. Stephen Curry (317)
3. Nikola Jokic (295)
4. Shai Gilgeous-Alexander (282)
5. Jalen Brunson (248)

These results were very good to see from a sanity check perspective, as the top 5 seems incredibly accurate. The list contains the top 4 candidates for MVP this season (Doncic, Jokic, SGA, Brunson) as well as the widely-considered greatest shooter of all time in Curry. We can see that front offices would be wise to examine this statistic if they are looking to increase their scoring juice in the offseason. Potential targets of note that performed well in POE were Collin Sexton (12th, top 11 were all All-Stars) and Payton Pritchard (31st despite limited minutes compared to most).

4.3.2 RAPM Model - Expected Points

As discussed earlier in 3.2, I wanted to compare players based on their on-court impact on total expected points. I believe this statistic could help front-offices see how certain players "gravity" or off-ball impact offensively helps create better shots for their teammates, as well as which top defenders do the best job at keeping the opposing team away from high-value areas of the court. Adjusting on a per-minute basis can also inform teams as to which players perhaps deserve more playing time. Below are the top 10 players in the NBA in terms of their coefficient in the RAPM model:

1. Giannis Antetokounmpo (0.0840)
2. Franz Wagner (0.0638)
3. Amen Thompson (0.0612)
4. Ausar Thompson (0.0590)
5. Josh Hart (0.0545)
6. Andre Drummond (0.0539)
7. Jalen Brunson (0.0533)
8. Draymond Green (0.0516)
9. Jonathan Isaac (0.0477)
10. Dean Wade (0.0477)

The results from the RAPM model are quite interesting. At the top, we have the do-it-all superstar Giannis and another rising star in Franz Wagner, who both have outsized impacts on their respective teams. We can see that the model seems to value a certain kind of player that often gets overlooked by more traditional statistics: the "glue guys". Guys like Josh Hart, Draymond Green, and Jonathan Isaac all play with tremendous energy and impact defensively, but aren't very impressive scorers and don't see a ton of the ball on offense. Box-score stats like PER and the like often tend to underrate these players as defensive statistics continue to lag behind, but we can see through the RAPM model that these players have an impact in creating better-quality looks for their team (perhaps through cutting, screen-setting, and/or offensive rebounding), as well as their defensive chops preventing the opposing team from getting to their spots offensively. We also see the value of the floor-spacing shooting archetype with Dean Wade, who helps the Cavaliers spread defenses out with his long-range accuracy.

5 Discussion

Overall, the concept of shot quality in basketball, much like in other sports soccer and hockey, is a multifaceted one. This project aimed to produce a shot quality metric and apply it to multiple aspects of NBA analysis, including team evaluation, playoff predictions, and individual player assessment.

The metric itself, computed utilizing a logistic regression model, gives us a baseline estimate of the chance of a given shot going in. This can be useful to know for coaches in particular, as they can cater their offensive and defensive schemes and play designs to maximize their expected points and minimize their opponent's. As expected, the model estimates that dunks (the baseline shot type) are the most effective shot type, followed by layups, jump shots, and then hook shots. NBA teams have been trending away from bask-to-the-basket post play for years now, and the numbers here back that assessment up.

When looking at team success, the results revealed interesting nuances. Utilizing raw expected points and expected points allowed didn't do a very good job in picking out the best teams in the league. The popular slogan remains "it's a make-or-miss league," and we can see here that having shot-makers (and therefore scoring more points over expectation) was more important when predicting wins than simply creating better looks.

In terms of predictive power, especially when looking at this year's playoffs, the results were definitely mixed. While traditional metrics like win percentage and point differential remain the cream of the crop when predicting team success, I believe points over expectation can offer valuable context into efficiency and talent as it relates to winning. However, it's essential to acknowledge the limitations of this approach, particularly in predicting playoff outcomes, as evidenced by the discrepancies between model predictions and actual results.

At the individual player level, points over expectation provided a insightful measure of scoring efficiency and shot-making ability. The top-ranked players, including Luka Doncic, Stephen Curry, and Nikola Jokic, are widely known for their offensive prowess, validating the efficacy of the shot quality metric in identifying the top players. The regularized adjusted plus-minus model also highlighted the contributions of players beyond traditional scoring metrics, emphasizing the value of defensive prowess, off-ball movement, and floor spacing in enhancing team performance. The main discrepancy in our earlier playoff model, the Phoenix Suns, would perhaps have fared much better in the postseason had they focused more on acquiring these "glue guys" and striking a better balance between shot-making and intangibles. Front-office executives at the NBA level should make sure to take a holistic view when evaluating players, as both Points Over Expectation and the RAPM model are simply tools to better aid decision-making.

In future research, incorporating additional variables such as player tracking data or opponent strength could help improve the basic shot quality model and provide a better estimate of "efficient" shots in basketball. Ultimately, metrics such as shot quality are only one piece of the puzzle. Basketball is a fluid and dynamic sport with many constantly changing variables, and so any attempt to quantify aspects of the sport will have its limitations. But by combining advanced analytics and mathematical methods such as the ones discussed in this paper with coaches and players expertise, teams can gain a leg up on the competition and make better decisions to increase their chances of being the last team standing come June.