Paper

Vedanth Pothina

Abstract

This project studies the success of NBA players drafted between the years 1995-2015 by showing their career success relative to their draft position. Averages were compiled for each of the players in the dataset using a variety of data points including career points, rebounds, assists, steals, blocks, turnovers, three point percentage, free throw percentage and field goal percentage. In order to evaluate pure performance without any draft position weight attached, a formula was created called PerformanceScore in which it incorporates all of the statistics we created into a formula that encompasses the entirety of a player's career. For each position (G,F,C), the stats are weighted differently using a regularized linear regression model in the PerformanceScore formula. Next, the ExpectedScore metric was produced to represent the predicted PerformanceScore for a player based on their draft position. Finally, the FinalScore was created by combining both PerformanceScore and ExpectedScore to get a ranking of every single player drafted from the year 1995-2015's career success relative to their draft position. This analysis provides an insight into under and over achievers throughout this 20 year span while also being able to provide General Managers insights and trends for future draft selections.

Introduction

NBA teams and fans are in dire need of a singular metric to determine whether a player they drafted underachieved or overachieved relative to their draft position. Understanding this will help them as they go into future drafts as they will see similarities in players from the past they drafted and future players they may want to draft. The FinalScore formula gives these teams just that as it is an unbiased formula to show their success relative to their pick and position.

This study has numerous objectives. It first examines how players drafted between 1995-2015 performed relative to their expectations based on their draft position. In theory, the players that have been drafted in the lottery and the first round should be the best in the entire draft. This is almost never the case. This paper examines the players who are able to overcome the adversity of being drafted in the later-rounds and provide that same superstar effect as the lottery players while simultaneously looking for the lottery players who never reached that height. Using the analysis done during this 20 year period, it will become clear without bias who the true underperformers and overperformers were from draft night.

Questions that surround the topic of expected draft value are questions such as which position provides the best value? Which draft group returns the most value by pick, lottery, mid-first or second round? On prior research, there were not many studies surrounding expected value for pick and rankings, especially not in the time period of 1995-2015. The research question for this project is how can we measure how NBA players drafted perform relative to their expected value by draft pick?

Methodology

To address the question above, the dataset is filtered down by year. For the purposes of this project, the dataset is purely NBA players drafted between the years 1995-2015 which, according to *The three Eras of the NBA regular seasons: Historical trend and success factors* by Joao Vitor Rocha da Silva and Paulo

Canas Rodrigues is the true start of the "transitional-era" of basketball. Player performance independent of draft pick was quantified using a metric called PerformanceScore, a weighted formula composed of the key offensive and defensive statistics, made using a ridge regression optimization formula. Different weights in the formulas were developed for guards, centers and forwards to account for position-specific roles on the court. The metrics used to evaluate PerformanceScore are points, assists, rebounds, turnovers, steals, blocks, free throw percentage, three point percentage and field goal percentage.

An ExpectedScore was then derived by evaluating the mean PerformanceScore per draft pick, showing the expected PerformanceScore associated with that selection in the draft. Finally, a FinalScore is calculated as the distance from the mean per pick or PerformanceScore-ExpectedScore, quantifying how much a player under or over performed relative to their draft pick.

The data used from this project was collected using basketball-reference.com, a popular basketball statistics site where the players in the datasets career-wide statistics were collected. Basketball Reference returned the Points, Rebounds, Assists, FG Percentage, Three Point Percentage, Free Throw Percentage, Pick and Name. In addition to basketball-reference, hoopR's API was used to find each player's career steals, blocks and turnovers.

For this project, the sample size consisted of all NBA players drafted from the 1995 NBA Draft till the 2015 NBA Draft, providing a total of 1,216 players without the inclusion/exclusion criteria. Due to roughly 5% of NBA players per draft class never playing a game and about 18% of players not providing statistical importance to the dataset, a filter was applied to remove outliers whose performance was inflated or deflated by limited opportunities in their career. All players in the final leaderboard played at least 2 years in the NBA which is the amount needed to put up statistics that will give a player a viable PerformanceScore.

Some of the key variables in this project were all of the major statistics that went into creating the formulas and the PerformanceScore, ExpectedScore and FinalScore. Players with missing data were either due to the fact that they hadn't played a game and were filtered out or because of the hoopR API. The hoopR API had a different set of names for players with accents in their names than basketball reference so those players' steals, turnovers and blocks were entered in manually.

Players with missing data in the dataset were excluded for 2 primary reasons: either they had been drafted and had never played a game in their career or their statistics were unavailable due to inconsistencies in either hoopR's API or basketball-reference's website. Specifically, there were numerous issues around players with special characters in their name as basketball-reference and hoopR handled name-formatting differently. These happened particularly for players with non-ASCII or accented characters in their names. In these special cases, data was entered manually to ensure data completeness in the dataset.

An extremely critical component in evaluating a player's importance to their team and to their statistics is their availability, which represents how often a player is active on the court. In the modern NBA, availability is one of the most important statistics in the game. A player like Mikal Bridges was known as a "role" player but helped the Suns make a deep run to the NBA finals due to his insane availability of playing every single game that season. Per-game statistics can be inflated if a player misses games consistently due to injury or personal issues. To account for this, an AvailabilityScore metric was created, a ratio of the games played to the maximum possible amount of regular season games in their career. This measure normalizes PerformanceScore by incorporating reliability, ensuring that the FinalScore also represents how valuable a player is to their team through their sustained presence on the court.

For each position in basketball, the appropriate PerformanceScore formula would be different as players in different positions are drafted for different roles on the team ex. (the guard's role is to run the offense while the center's role is to collect rebounds and score in the paint.) Using that idea, I selected the key offensive and defensive performance metrics, Points, Assists, Rebounds, Field-Goal Percentage, Three-Pointers Made, Turnovers, Steals, and Blocks and applied a ridge regression optimization model to determine the relative weights of each variable. These optimized coefficients were then used to construct a composite formula that produced a single clutch score for every player. They are as follows. (Put the optimization model result here and put an explanation of the model result) (issues w/ optimization model will talk about in next session) The formula for **Guards** is:

$$\text{PerformanceScore}_G = 0.301392942 \times PTS + 0.131585761 \times TRB + 0.245252914 \times AST + 0.071600908 \times STL +$$
$$0.088402778 \times BLK - 0.110155566 \times TOV + 0.003795444 \times FG\% + 0.029598577 \times FT\% + 0.018215110 \times 3P\% +$$
$$0.015 \times AvailabilityScore$$
$$(1)$$

The formula for **Forwards** is:

$$\text{PerformanceScore}_F = 0.278032221 \times PTS + 0.035431424 \times TRB + 0.363261918 \times AST + 0.003023791 \times STL + 0.14394952$$
$$(2)$$

The formula for **Centers** is:

$$\text{PerformanceScore}_C = 0.166423233 \times PTS + 0.099124073 \times TRB + 0.377695524 \times AST + 0.006508985 \times STL + 0.09341567$$
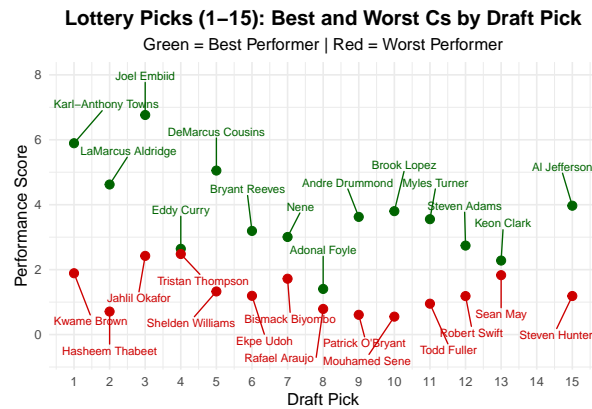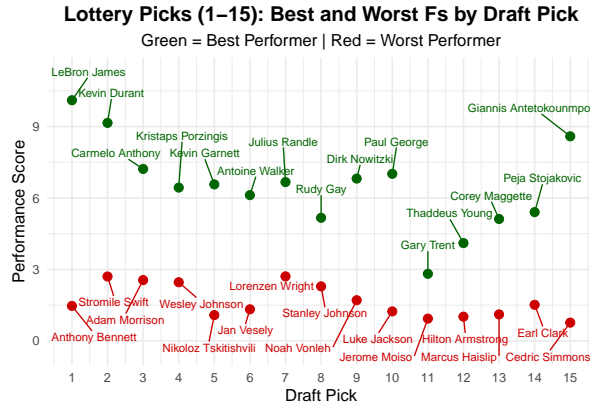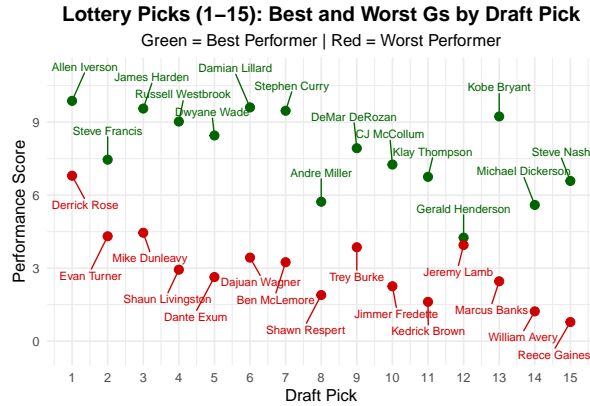$$(3)$$

To calculate the ExpectedScore, the average PerformanceScore per pick is found. For example, all players regardless of position at the 17th pick will have the same ExpectedScore. The ExpectedScore also shows that the average player per pick, even in the lottery, is not some sort of superstar but most of the time, a high end role player. Finally, FinalScore is calculated by PerformanceScore-ExpectedScore, essentially finding how far away from the mean per pick the player is.
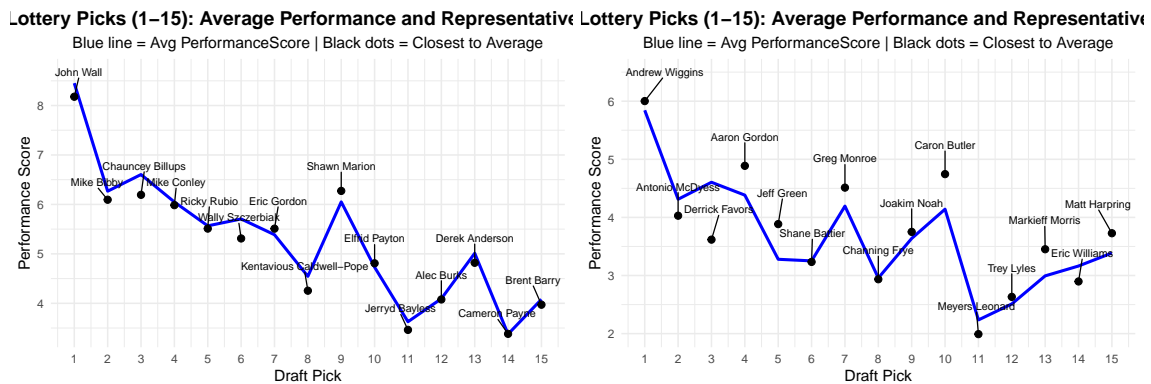
Results

After the ridge regression optimization model to create weights and apply the PerformanceScore, ExpectedScore and FinalScore to all the players in our dataset, a series of leaderboards filtered by position were returned.(Leaderboards will be in Appendix). Figure 1 shows the average PerformanceScore per pick. When looking at Performance per Pick, the mean PerformanceScore/ExpectedScore is not a linear decline with variability as the last overall 60th pick has the highest ExpectedScore after the top 10 picks. The graph also shows the low variance of the lottery, as most of the top 15 picks are "safe" while simultaneously showing that the late first to second round picks have a lower average return but a greater variance of pick. Outcomes in this latter part of the draft are spiky, not steady as the median value is low but outliers(Nikola Jokic, Manu Ginobili) pull the mean up.

Average Player Performance by Draft Pick

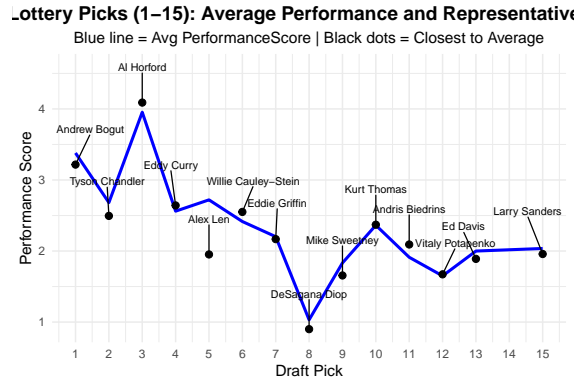## Mean PerformanceScore aggregated across all positions
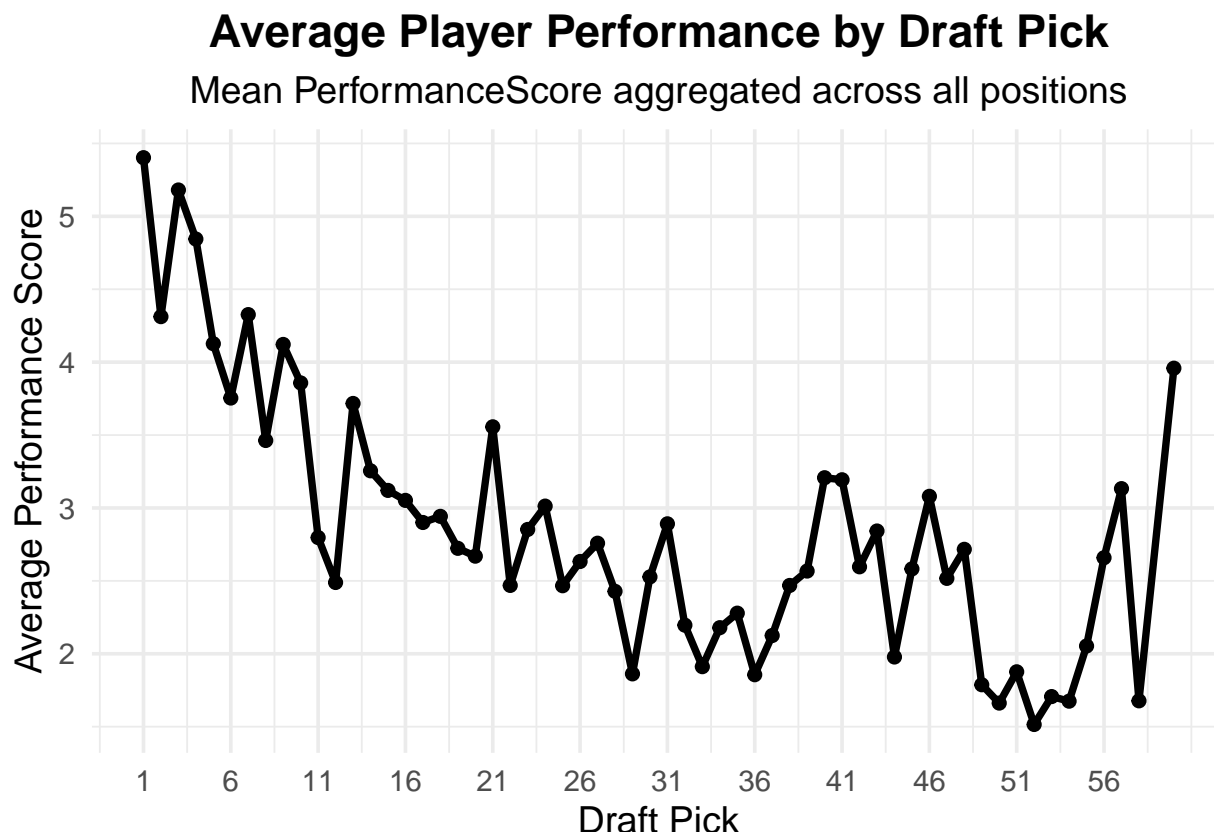
When drafting, a general manager expects to get the best players he can in the lottery(1st to 15th pick). When analyzing these picks by position, clear positional trends emerge. Figures 2a, 2b and 2c show the best and worst value per pick among the top 15 picks in the draft per position. Among forwards, there is the steepest drop off after the top picks(Kevin Durant and Lebron James) but also show a late-lottery success story in Giannis Antetokounmpo. When focusing on the guards, an unique pattern emerges when looking at the PerformanceScore variance. As expected, the top performers and superstars in the league(Stephen Curry, Kevin Durant and Kobe Bryant) validate the model's ability to identify elite talent, the lower end or players in the "red" among the lottery picks are not poor performers at all and show just how good guards that have been selected in the lottery are. In fact, players like Derrick Rose and Shaun Livingston were considered to be stars in the league. In comparison to both forwards and centers, the lower end of the guards are well above the average range. In terms of centers, most of the centers in the final top 10 were not drafted in the lottery. In Figure 3C, there is low quality of performance among all the centers in the "green", with many of them having extremely close PerformanceScores to their counterparts in the red, showing that there is no real point of taking a Center in the lottery. For example, Tristan Thompson represents the worst 4th overall pick that's a center with a PerformanceScore of 2.480817 while Eddy Curry represents the best having a PerformanceScore of 2.641428, only 0.160611 over Thompson.

Lottery Picks (1–15): Best and Worst Gs by Draft Pick
Green = Best Performer | Red = Worst Performer

Lottery Picks (1–15): Best and Worst Fs by Draft Pick
Green = Best Performer | Red = Worst Performer

Lottery Picks (1–15): Best and Worst Cs by Draft Pick
Green = Best Performer | Red = Worst Performer

After looking at the best and worst players per lottery pick, the natural question arises: When drafting in the lottery, what type of players should one expect per pick? Figures 3a, 3b and 3c examine the player closest to the average Performance Score for their pick among the top 15 picks in the draft. The blue line visualization demonstrates how average player performance declines with draft pick, but the curve is not perfectly linear in all 3 positions. The black dots above and below the line represent the players drafted at the pick shown that are closest to the average. These players show draft analysts and general managers what their baseline is when drafting in the lottery and how they should go about strategizing around that. Mid-lottery players over exceeded expectations at times, showing the natural variability of draft outcomes.

Lottery Picks (1–15): Average Performance and Representative
Blue line = Avg PerformanceScore | Black dots = Closest to Average

Lottery Picks (1–15): Average Performance and Representative
Blue line = Avg PerformanceScore | Black dots = Closest to Average

Lottery Picks (1–15): Average Performance and Representative
Blue line = Avg PerformanceScore | Black dots = Closest to Average

When looking at the average FinalScore for every team in Figure 4, the San Antonio Spurs come out on top having an average FinalScore of approx 0.91656023 over 0.5 over the next best team of the Oklahoma City Thunder. From 1995-2015, the Spurs were a dynasty, having great players such as Tim Duncan and Manu Ginobili. This analysis just reinforces their front office's role in their success. On the other hand, the Dallas Mavericks had the worst average having an average Final Score of approximately -0.870. Seeing as the Mavericks won a championship in 2011, this is extremely interesting. Upon further analysis, we can see that they traded for their star player in Dirk Nowitzki who was drafted by the Bucks on draft night, showing how they obtained their success.(Note, as shown in the previous example, we can not attribute a player's success to the team that drafted them as the Milwaukee Bucks played no role in the success of Dirk Nowitzki's career so take this analysis with a grain of salt). This analysis benefits general managers in 2 ways, they first can analyze team success using this graph and try to model their drafting strategies around the most successful drafting teams. They can also use this to hire and recruit scouts or other people in the front office that have been successful when drafting championship-level players and teams in the past. While this graph doesn't show the people in the front office on these successful teams, you can use a website like Linkedin to find and recruit these scouts.

Average FinalScore by Team
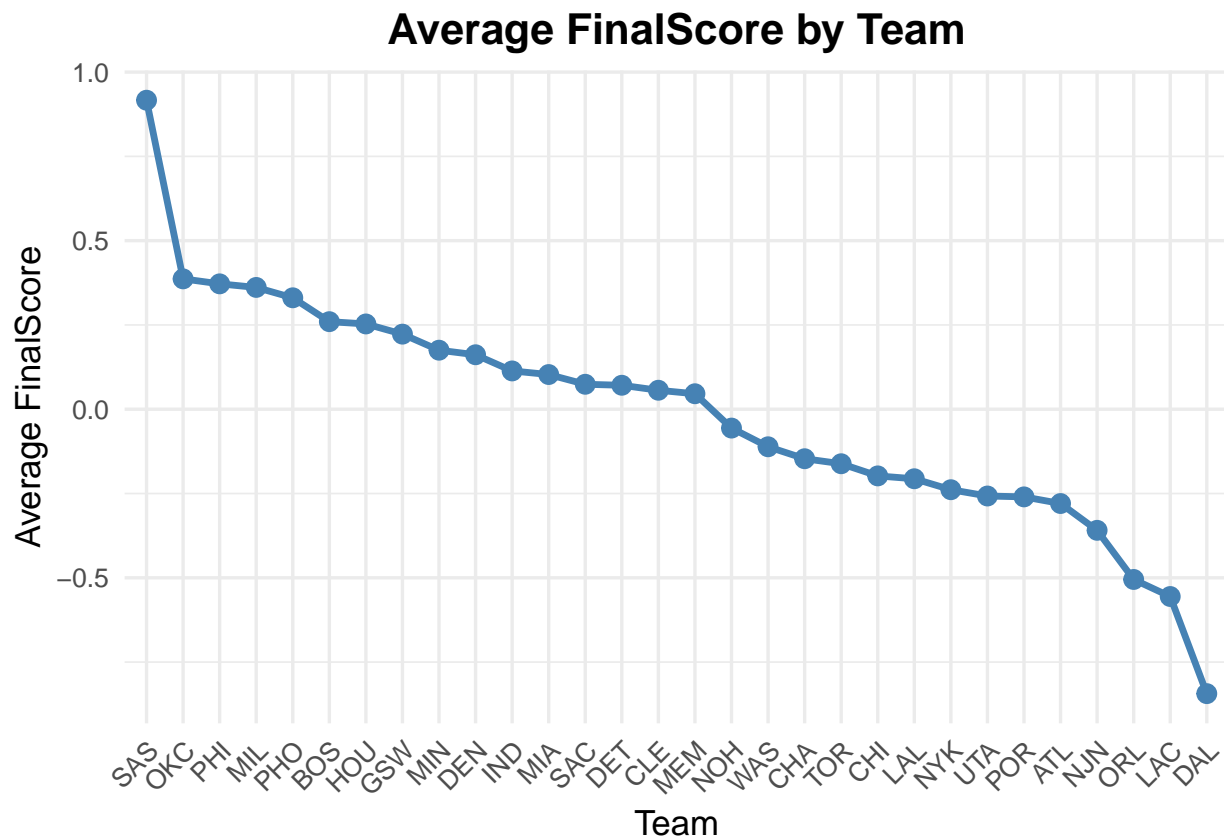
Conclusion

This study developed a data-driven framework to analyze the success of NBA players drafted between 1995-2015 by quantifying how well players performed relative to other players at their same draft position. Through PerformanceScore, ExpectedScore and FinalScore, this analysis was able to capture both absolute and relative career value for each player in the dataset. By weighting all statistics using a ridge regression optimization model, a dataset that accounted for positional differences was returned.

The results show the true positional trends in the lottery. Guards exhibited the narrowest range of outcomes, showing that they have the least overall draft risk when choosing who to select. Forwards showed the most variance producing top caliber players in the early lottery picks with a big dip in the middle and a resurgence with late lottery value picks as well. Centers showed that they should not be taken in the lottery as most of the best centers in terms of PerformanceScore and FinalScore were taken in the later picks with a low quality of performance in centers in the lottery.

In this project, others may have doubts around the PerformanceScore formula which is why a webapp in which you can remake the formulas to fit what you think the ideal guard, forward and center should be scoring. The formulas in this project is purely a standard in which the average center, guard and forward can be assessed. Something else that could be used to find different weights is to use a different optimization model such as Lasso or Elastic Net regression. In attempting to create the best FinalScore formula, z-score was experimented with, before FinalScore was just a straight deviation from the mean. When using z-score, the results were very skewed and the leaderboards did not accomplish the project goal due to them blowing up outliers and having an overemphasis on draft position rather than a combination of both pick and performance.

Future work in this topic could include predictive validity, using this analysis and finding similarities in players in upcoming drafts or players that have just been drafted to predict how well they will do relative

to their draft position in the future.

This analysis would be most applicable towards the draft analysts and key basketball operations decision makers, especially in the media. These draft analysts can use the proposed methodology see the quantifiable "floors and ceilings" each player has based on pick while also seeing what to expect from each pick. These insights can then be used to do the most critical thing a front office can, making more data-driven and efficient decisions in the future to better their team.