Nicholas Andrews

Byeong-Hak Choe

DANL 399

Dec 17, 2023

**The Greatest NBA Playoff Lineup**

Introduction

Who would you pick if you wanted to create an all-time starting 5 to win the NBA title? Would you pick the best scorers? After all, a team simply needs to score as many points as possible to win. Or would you pick the best defenders? If you prevent the opposing team from scoring, your team will probably win. You may want the most unique players. Having un-guardable moves like Steph Curry’s three-point shot, or Kareem’s sky hook. These conversations are referred to as “Barbershop” talk. A fun, sometimes heated debate about certain sports takes. As fun as these talks are, they tend to be merely subjective. Even when individuals bring up counting statistics (typically “per game” statistics), these statistics do not paint an objective view of the topic at hand. For example, if a player plays in a slower-paced era but impacts the team's offense tremendously, their “Points Per Game” may be less than an inferior player who plays in a fast-paced offense.

This data analysis aims to fix this problem. Advanced metrics do the best job at measuring a player's impact regardless of uncontrollable variables like the pace of play or what era a player played in. Using pure advanced metrics, and championship titles, the best possible NBA playoff starting lineup can be created. But again, what kind of team is the best? This dataset uses advanced analytics that show an impact on winning. The advanced analytics used do not just measure how good of a scorer, or defender a player is, just how much a player impacts winning. Regardless, a player should be judged by how much you impact winning, not how much you can stuff a box score. This data analysis strides to find the best player at each position who impacts winning the most in the playoffs.

Methodology

**Variable Description**

This dataset uses variables that only account for a player's playoff statistics. A player should not be judged by how they do in the regular season. The goal of a player should only be to win a championship. So, if a player gets amazing regular season statistics, but is a lesser version of themselves in the playoffs, the regular season statistics must be ignored, Alternatively, if a player plays significantly better in the playoffs, this should help their case for being a great player.

Additionally, this dataset aims to use more “Per Season” statistics. This is done to limit the impact longevity has on the results. Although longevity is important, the length of a player's career should not weigh the same as the quality of a player's career. For example, if a player is a great player but decides to play past their peak to a point where their play hurts their team, longevity statistics count these less productive seasons as a positive. Using “Per Season” statistics, this can be accounted for and treat the less productive years as a negative for the players' claim to greatness. The variables used are as follows:

*'PER\_playoffs': Player Efficiency Rating for a player in the playoffs.*

*'Offensive\_win\_shares\_playoffs\_per\_season': Offensive win shares in the playoffs per season.*

*'Offensive\_rating\_playoffs': Offensive of a player in the playoffs.*

*'VORP\_playoffs\_per\_season': Value over replacement for a player in the playoffs per season.*

*'Box\_plus\_minus\_playoffs': Playoff Box Plus Minus for a player in the playoffs for their career.*

*'Number\_of\_Championships': Number of Championships a player wins in their career.*

*'Win\_shares\_per\_48\_playoffs': Win Shares for every 48 minutes in the playoffs for their career.*

*'Defensive\_win\_shares\_playoffs\_per\_season': Defensive win shares in the playoffs per season.*

These variables best rate a player's impact on winning. Regardless of the pace of play, the surrounding cast on a player's team, or the era they played in, these statistics objectively grade a player's impact on winning in the playoffs.

**Weights**

Although these statistics are great, some of them do a better job than others at rating players. For example, many analysts think Value Over Replacement Player is a better statistic than Player Efficiency Rating. To account for this, some variables were weighted differently than others. The weights are as follows:

*PER\_playoffs = 7,*

*Offensive\_win\_shares\_playoffs\_per\_season = 8,*

*Offensive\_rating\_playoffs = 10,*

*VORP\_playoffs\_per\_season = 11,*

*Box\_plus\_minus\_playoffs = 12,*

*Number\_of\_Championships = 14,*

*Win\_shares\_per\_48\_playoffs = 13,*

*Defensive\_win\_shares\_playoffs\_per\_season = 8*

**Valuation**

Before weighting the variables, the data was standardized to account for the differing values. This made the value of each variable have equal means and standard deviations so the results are not skewed by values that are measured differently. To calculate the composite score, standardized variables were multiplied by the corresponding weights, and then summed.

After the proper weights were applied, the variables were then used to generate a composite score. This composite score is what determines the value of the player's impact on winning in the playoffs.

**PCA Regression**

Beyond the question of who the best players are, it is worth exploring what variables impact winning the most. To find this out, a PCA regression was run on the dataset. Since the data has many variables, it felt appropriate to use a PCA model to reduce the dimensionality of the data.

Results

**Composite Score**

To reiterate, when analyzing the best playoff performers from the dataset we look at the composite score. The higher the composite score, the better a player is. Overall, Michael Jordan has the best composite score with a value of 175. The best possible lineup consists of; Magic Johnson at PG, Micheal Jordan at SG, Lebron James at SF, Tim Duncan at PF, and Nikola Jokic at C. Magic Johnson, Micheal Jordan, and Lebron have large gaps over the next player in their respective positions. Tim Duncan has a slight advantage over Giannis Antetokounmpo, and Nikola Jokic has a slight advantage over Kareem Abdul-Jabaar. All these players are highly praised for their performances, and most people would agree with this lineup… except for the Center position. NBA history is rich with great playoff Centers. With names like Kareem, Shaquille O’Neil, Hakeem Olajuwon, and many more, it is a bit surprising to see a current player like Nikola Jokic take the top Center position. I would argue this speaks to the greatness of Jokic. Jokic is well known for his advanced analytics, consistently at the top of these statistics each season. It is also well known that his impact goes well beyond typical counting statistics. I would not argue that Jokic is the best playoff Center of all time, it is too early in his career to make that claim. I will however argue that if he keeps this production at this level, he will be the greatest Center of all time in the playoffs.

**PCA Regression**

The PCA regression revealed what variables impact composite scores the most. The most impactful predictors were Offensive Win Shares per Season in the Playoffs with a coefficient of 25.23, and VORP in the playoffs with a coefficient of 24.27. These coefficients essentially mean that if you increase Offensive Win Shares by 1, the composite score increases 25.23 times holding all other variables constant. The lowest variable coefficient was Win Shares per 48 with a coefficient of 13.73. It is worth noting that this variable still impacts the composite score by 13.73 times holding all other variables constant, which is still a high number. These coefficients suggest that all these variables impact composite scores by significant amounts. This makes sense because the composite score is derived from the data for the variables.

The adjusted R^2 has a value of .9972. This means that there is virtually no variance in the data derived from the variables. This indicates a high correlation between the variables and the composite score. It also indicates a high level of accuracy, which is an indication of a good model.

**Conclusion**

Given the high R^2 values and coefficients, it is worth paying attention to the regression’s correlations. All the variables used are highly correlated with composite scores. This means if you were a GM whose main goal is to win in the playoffs, the advanced metrics used in this data set must be paid attention to when building a team. Getting players the excel in Offensive win shares and VORP may produce more playoff success. It must be noted that players before the 1971-1972 season were not used because of the lack of data. This means players like Bill Russell and Wilt Chamberlain were not used, who could have shaken the results up given how much success both of them had.

For future research, I would recommend researching what systems produce the most success. Lots of titles are won because of a specific type of system, which means some offensive/defensive systems work better than others. If GM can get the right kind of players and the perfect playoff system, this team undoubtedly will have great playoff success.