Prepare & Visualize Data

The NBA plus minus has always been an interesting topic, as far as how it actually correlates with a player and their performance. In this assignment we will look at the 2017-18 first round playoff matchup between the New Orleans Pelican and the Portland Trail Blazers. The first analyses will be to look at the simple plus minus for both of these teams regular seasons. The major thing to realize for the simple model is it is adjusted to represent as if the player was on the court for 48 minutes.

The second model we are going to look at is the win probability plus minus. The win probability plus minus is similar to the simple plus minus except that win probability is in place of the scores. We will need to look at players on both of the teams in order to figure out more info about our models. At first glance it will be difficult with 82 games for each team, so we will need to find a way to average the data to make it easier to analyze.

Review Measurements & Justify Choices

The first choice we make is to analyze a players plus minus over 48 minutes. We will see if this is an accurate representation of how good a player is and how well they perform. The main measurement issue for the simple plus minus is that a players stats can be inflated or deflated if they only play a few minutes. The team could outscore an opponent by a good amount while they are on the court. For instance a team could be down by 20 points in the last minute of the game and let's say they outscore the opponent by 6 points. If you aggregate this over 48 minutes the

players simple plus minus will be extremely inflated. The opposite could also happen and the simple plus minus would be deflated.

In the second model we replace the scores with win probability. We then calculate the affect that players on Portland and New Orleans had while they were on the court. In this case a team with a five point lead at the beginning of the game will have a much lower win probability than a team with a five point lead with 30 seconds left in the game. This will probably be a better indicator than the simple plus minus model because the players who play very few minutes will only be in the game when the game is already decided. Therefore, they will not affect the win probability as much as it will stay around 0% or 100% depending on the team you are on.

Implementation & Programming

In order to implement the simple model we first only want to look at games with the New Orleans Pelicans or the Portland Trail Blazers. These are the only two teams we are analyzing so we eliminate the rest of the games. We look at the play by play for each one of these teams by game, and factor in how long each person is on the court. We then factor in how the score changes while each player is on the court. We take these factors and use the aggregate function in R to determine a players plus/minus. Then we convert the players plus minus for the game as if they had played 48 minutes. To calculate this we divide the amount of seconds in a game by the amount of seconds they played and multiplying by their plus minus for that particular game. The final step of implementation is to average each players simple plus minus for all the games.

The first step for implementation for the win probability adjusted model is to build a model that determines the away teams' chance of winning, depending on how much time is left in the game. We then used a logistics model that factored in the score difference and game

elapsed time and factored this in with the chance of the away team winning. This will give us the probability a player increases the team's chance of winning while they are on the court, while factoring in the current win % for that team.

Results & Variability

For the simple model with 82 games being played during the regular season it is a little difficult to truly analyze the results of the players for each game. We averaged the simple plus minus by player to get a better visualization in R. A couple things stick out when we do this. The main starters for the Pelicans throughout the year were Anthony Davis, Jrue Holiday, E'Twaun Moore, Rajon Rondo and Ian Clark. Anthony Davis had the highest adjusted plus minus per game at 5.44. The lowest of these was Rajon Rondo at -.98. For the Portland Trail Blazers the main starters were CJ McCollum, Damian Lillard, Jusuf Nurkic, Al Furuk Aminu, and Evan Turner. The highest of these for the simple plus minus was Damian Lillard at 6.71 and the lowest was Evan Turner at -1.42. All of these starters were fairly close to even which shows this may be a pretty good representation. Where we found inflation or deflation was with Georgios Papagiannis, Larry Drew II, and Josh Smith. Papagiannis played in 1 game and had an adjusted plus minus of +192. Smith played in 3 games with a -64 rating and Drew played in 7 with a -65 plus minus rating. This just shows that with minimal playing time a players adjusted plus minus can be greatly inflated or deflated.

Again for the win probability model it is a little difficult to tell the results with 82 games so we will average the results per player to get a better idea. Again let's first look at the main starters. Of the Pelicans starters the highest was Jrue Holiday at .03 and the lowest was Rajon Rondo who averaged at about even. For the Trail Blazers starters the highest was Al-Farouq Aminu at .0018 and the lowest was Damian Lillard at -.04. Again it does appear that the players

who play less still may be inflated or deflated as Georgios Papagiannis was the highest at .33 and played very little.

Recommendation, Results, & Exposition

For the simple plus minus it is pretty clear that it is a decent measure for players with a good amount of playing time. However it is a terrible measure for players who play very little as the number can easily be inflated or deflated. It is well known that Anthony Davis and Damian Lillard are All Star basketball players, so it would make since that they have a 5.4 and 6.7 respectively simple plus minus. One other thing to note is that a players simple plus minus can be greatly changed by the players around them. In this case both of these teams were good, but not great.

For the win probability plus minus model it seems okay, but I still think in order to get a good idea of how good a player is they need to play a good amount of time. In our model we saw Papagiannis had the change in win percentage while only play for a couple minutes in the year. This shows that for players who only play a few minutes this is probably not a good representation either. One main difference between the models was Damian Lillard had a very high rating in the first model, but was lowest of all the Trail Blazers starters in the second model. Lillard is known as the best player on Portland, so this would indicate that the first model may be a little better. However, what is not factored in is all the teammates that are around Lillard and how they play has a direct reflect on both of these models.

If we look at the Real Plus Minus on ESPN we see that it is calculated by 100 offensive and defensive possessions. This also takes into account opponents. Anthony Davis was the Pelicans top player with an RPM of 5.55. This ranked 7th in the NBA. The Trail Blazer's top

player was Damian Lillard. He ranked 14th overall with a 4.90 RPM. This shows that maybe the simple plus minus could be a pretty good indicator if a player has enough minutes as Davis and Lillard were the highest in that model as well.

Playoff Model & Results

For this model we will look at our original simple plus minus model for the Portland Trail Blazers and New Orleans Pelicans. The actual results for the games were the Pelicans won game one 97-95, game two 111-102, game three 119-102, and game four 131-123 for a 4-0 series sweep. We will now look at our model and predict the results of these games. Instead of using the regular season data we will use the playoff play by play. If we sum up the results by each game we can analyze off of this model who would win in our prediction. The average in the first game is Portland is at 202 and New Orleans is at -50, for a Portland win. For the second game New Orleans is at -3.5 and Portland is at -574 for a New Orleans win. For the third game New Orleans is at -146 and Portland is at -49, for a Portland win, and for the fourth game New Orleans is at 46.5 and Portland is at -127 for a New Orleans win. So according to our model the series should have been tied at two with Portland winning game 1 & 3 and New Orleans winning game 2 & 4. However, New Orleans won all four games. The reason why this model is off is the adjusted plus minus aggregates it per 48 minutes. So for instance Wade Baldwin IV played 17 seconds in game 2 and had a plus/minus of -3 while he is on the court. However, if you convert that into 48 minutes it would be -508. Another method is take the average by team, similar to what we had done in the regular season, with the exception that we were averaging by team instead of player. We get the same predicted results with Portland winning games 1 & 3 and New Orleans winning games 2 & 4. We can also do what we did in the regular season model and run the average by player. The main point from these results are the majority of the numbers are far

away from 0. This is because we have a small sample size. We only have 4 games compared to 82 in a regular season. These numbers also get inflated or deflated based off of the sample size.

I would tell management that the simple plus minus and the win probability plus minus would be an okay measure for anyone who plays a good amount of minutes. However, for players who play little it becomes inflated and deflated. I think a plus minus of while the player is on the court would actually be a better indicator than the simple plus minus per 48 minutes.