

The 2019-20 NBA Season: What Could Have Been

Ryan Kohanski

Rithvik Saravanan

Howard Yong

May 11, 2020

Abstract

The COVID-19 pandemic has affected our society in various ways and has changed numerous events on schedule for 2020. One such event that we were looking forward to was the end of the 2019-20 NBA season as well as the 2020 NBA Playoffs. Since a large portion of the 2019-20 NBA regular season games have already been played, we utilized the data collected from these games to run regression prediction models and calculate Elo ratings for each team in order to predict the standings for 2019-20 as well as the matchups and results of the Playoffs and the NBA season awards. Using the Elo rating predictions by matchup, we predicted that the Milwaukee Bucks ended with the highest seeding (standing) at the end of the season with the Los Angeles Lakers coming in second. Running our analysis by predicting the 7-game playoff series matchups, we predicted that the Western Conference Finals matchup will be between the #1 seed Los Angeles Lakers and the #2 seed Los Angeles Clippers and the Eastern Conference Finals matchup will be between the #1 seed Milwaukee Bucks and the #2 seed Toronto Raptors. We found that running these simulations predicts the NBA Finals matchup between the Los Angeles Lakers and the Milwaukee Bucks with the Los Angeles Lakers ultimately claiming the Larry O'Brien NBA Championship Trophy. Additionally, we used several types of regression models to best predict end-of-season statistics for every player. We ultimately used a logistic regression model to predict the end-of-season statistics and leaders for each of the major categories and found that our predictions indicate that Giannis Antetokounmpo will claim both the NBA Most Valuable Player (MVP) award as well as the Defensive Player of the Year (DPOY) award. According to our prediction model, this will mark only the third time in NBA history that a player will win MVP and DPOY in the same season with the previous two players being basketball legends Michael Jordan and Hakeem Olajuwon.

Introduction

Due to the widespread impact of the COVID-19 pandemic throughout the world, almost every company, organization, and public event has canceled or suspended any activities that involve interpersonal contact for the foreseeable future. Many of these activities are moving to a virtual format if possible, but several others have been forced to shut down.

As avid sports fans, the absence of the major sporting events during this time has hit us and many others around the world especially hard [1]. Some of the events that we particularly were looking forward to include the NBA, NCAA March Madness tournament, MLB, and the 2020 Summer Olympics.

In our curiosity, we decided to utilize this opportunity to exercise our data science and modeling skills in order to predict what could have been. Specifically, we focused on the NBA and the NBA Playoffs. Since the 2019-20 NBA season was suspended approximately one month prior to the end of the regular season (and the beginning of the Playoffs), we used the 2019-20 season data accumulated from the games played before the suspension to predict how the season and the Playoffs would have ended had everything gone according to schedule.

In this analysis, we will examine data from the 2019-20 NBA season as well as some data from previous NBA seasons in order to draw some meaningful conclusions about the remainder of the 2019-20 NBA season including the final season standings, playoff matchups, championship winner, and season award winners.

Methods

Predicting the 2019-20 NBA Season Standings

Since we missed one of the most exciting times of the year (the NBA Playoffs & Finals), we made some predictions on how the rest of the season might have played out using a popular methodology referred to as the Elo ratings system [2]. This tool, created by Hungarian-American physics professor Arpad Elo, was originally designed to rate chess players, but is now used for all sorts of competitions ranging anywhere from sports to video games. This is a methodology that FiveThirtyEight and many other popular sports analysts take advantage of due to its simplicity and effectiveness [3].

These ratings depend only on the final score of each game as well as where it was played (home-court advantage). In other words, this system is built on a Win/Loss basis. We will be analyzing the 2018-19 NBA Season in its entirety to validate its performance, then we will apply it to the 2019-20 regular season in order to predict the matchups for the Playoffs and the Finals and ultimately the NBA Champions. For this project, we retrieved several types of data sources including game-by-game scores and schedules for several seasons from Basketball-Reference.com [4].

How does Elo work?

The long-run average for an Elo score in the NBA sits around 1500. An Elo of 1500 means that the teams performance would be normally distributed around an average of 1500 with the chance of performing better or worse. For more detail, Figure 5 (Appendix) shows what an Elo rating tells us about a team and how it can convey the teams overall season record. A higher Elo rating indicates that the team has a high win-loss ratio and is more likely to play deeper into the season.

The formula for Elo below shows how the probability of one team beating another is calculated using the ratings. When Player A competes in a match against Player B , Player A has an expected outcome (probability or score) for Team A ($E[A]$) where R_A is the rating for Team A and R_B is the rating for Team B . The expected outcome for Team A ($E[A]$) can be calculated by the formula below:

$$E[A] = \frac{1}{1 + 10^{\frac{(R_B - R_A)}{400}}}$$

The same calculation ($E[B]$) has to be done for Player B , but with R_A (current rating A) and R_B (current rating B) swapped so that $E[A] + E[B] = 1$. Once the match is played and S_A (actual outcome or score for Team A) and S_B (actual outcome or score for Team B) are determined, R'_A (the new rating for A) and R'_B (the new rating for B) are calculated with the formula below:

$$R'_A = R_A + K(S_A - E[A])$$

The S value in our case would either be 1 for a win, or 0 for a loss. This is because there are no ties in the NBA.

In this equation, K is an optimization constant that usually takes different values according the sport and the amount of games available. In other words, this value is the maximum amount by which a score can change in one match. If K is set too high, the ratings will jump around too much; if K is set too low, Elo will take too long to recognize important changes in team quality. Determining the right value of K is an entirely different and more complicated topic, so for this experiment we will be using $K = 20$, the optimal K for the NBA determined by FiveThirtyEight [3]. This is higher than most other sports and can likely be attributed to the fact that the NBA plays more games (81 games per team) and is subject to relatively little randomness.

Home-court advantage is set as equivalent to 100 Elo rating points. One hundred Elo points is equivalent to about 3.5 NBA points, so it can also be interpreted as the home team being favored by 3 to 4 points if the teams were otherwise evenly matched (obviously this value fluctuates from season to season). Since every team plays about half of their games at home and the other half away, a change in the home-court advantage value does not produce a significant difference in the ratings, but is still an important factor to consider.

Elo strikes a nice balance between ratings systems that account for margin of victory and those that do not. While teams always gain Elo points after wins and lose Elo points after losses, they also gain or lose more with larger margins of victory.

This works by assigning a multiplier to each game based on the final score and dividing it by a team's projected margin of victory conditional upon having won the game. For instance, the Golden State Warriors' 4-point margin over the Houston Rockets in Game 1 of the 2018-19 Western Conference finals was lower than Elo would expect for a Warriors win. So the Warriors gain Elo points, but not as many as if they'd won by a larger margin. The formula accounts for diminishing returns; going from a 5-point win to a 10-point win matters more than going from a 25-point win to a 30-point win. For the exact formula, see the footnotes.

Instead of resetting each team's rating when a new season begins, Elo carries over a portion of a team's rating from one season to the next. This is to account for any momentum that a team may build from season-to-season (i.e. sports dynasties). In NBA ratings, three-quarters of the previous score are kept. The high fraction reflects the fact that NBA teams are more consistent from year to year. For example, the Miami Heat ended the 2012-13 NBA season with an Elo rating of 1754. The team's Elo rating for the start of the 2013-14 season is calculated as follows:

$$(0.75 * 1754) + (0.25 * 1500) = 1692$$

Since this is a consistent method, we will also initialize the Elo scores for the 2019-20 NBA Season using the Elo scores from the 2018-19 season.

After incorporating a constant for home court advantage, our formula is as follows with $A = 100$ points (the value we previously determined represents a home-court advantage):

$$P(\text{Home team wins}) = \frac{1}{1 + 10^{-\frac{(H-R+A)}{400}}}$$

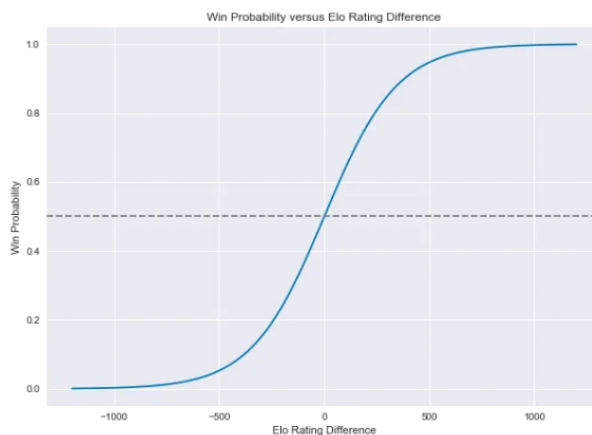


Figure 1: Logistic Function of Win Probability by Elo Rating Difference

In Figure 1, we see an example of a logistic function for win probability by Elo rating difference. The key takeaway from this figure is that the higher the difference between the teams' Elo ratings, the more likely that the team with the higher Elo rating is to win.

Calculating Elo Ratings

For our project scope, we decided to model our Elo rating system purely off of home-court advantage, wins, losses, and point-margin differentials. Of course, in the NBA, there is a wide array of different factors to consider, including roster changes, injuries, winning streaks, etc. These are all features that limit the capabilities of our model, but serve as potential targets of investigation for the future. As we mentioned before, one principal task of our project was to explore outcomes of the 2019-20 NBA season if the games were to continue. In order to do so, we first needed to develop Elo ratings for each team at their current state

of the 2019-20 season. Online resources report that the NBA has historically shown an average Elo rating of 1505, as the league is continually experiencing expansions and contractions that influence the center of the distribution of team ratings [3]. Accordingly, we followed this statistic in our Elo rating calculations. When simulating Elo rating calculations for each team throughout multiple seasons, we preemptively decided to carry over 75% of each team's Elo rating from the previous season, with the other 25% weight carried by the league average of 1505. This proportion was determined through historical league analyses of the NBA.

For the scope of our analysis, we decided to begin our analysis in the 2016-17 NBA season. We chose to start our calculations from this season because the season data was readily available and provided enough years to build a more-updated status of each team's current standing in the league. Once we had collected the data for the NBA seasons, we calculated the Elo ratings by the aforementioned methods. We iterated through the season by matchup, calculating and updating the Elo ratings for the teams playing simultaneously, and applied this to every season until 2019-20.

After calculating the Elo rating as mentioned earlier, we ended up with the Elo ratings of the NBA teams at the beginning of the 2019-20 season. We tracked their progress throughout this year's season and plotted them in the figure below. The plot is faceted by Conferences (East and West), which determines how many times a team plays other teams, as well as who they may contend against in their playoff runs.

Simulations and Predictions

After calculating the Elo ratings for each NBA team up until the last game they played (before the season was suspended), we used these ratings to then perform simulations matching up teams based on the rest of the season schedule as well as the post-season tournaments based on the top seeded teams from the regular season.

The Remaining 2019-2020 Regular Season

Since the 2019-20 season was suspended before the end of the regular season, there were still more games to be played (about a month remained in the regular season). Fortunately, we were able to scrape the remaining schedule for the regular season for every game that was set to take place.

The regular season would have originally lasted through mid-April. Since we have all the matchups, we can run these teams through a simulation to compete based on probability using their Elo ratings. This involved iterating over the dataset, retrieving each team's Elo rating to calculate their probability of winning/losing, and then pulling a random sample as the winner. After the winner was found, the Elo ratings were then updated to reflect the outcome of this matchup.

Repeating this procedure for every game in the remaining regular season schedule resulted in the final Elo scores for every team at the beginning of the Playoffs.

The 2019-2020 NBA Playoffs

In the NBA Playoffs, the top 8 teams in terms of win/loss record from each Conference (Eastern Conference & Western Conference) advance to the Playoffs. The teams are then seeded (ranked) based on these win/loss records, and matched accordingly for the playoffs. This means the #1 seed plays the #8 seed, the #2 seed plays the #7 seed, the #3 seed plays the #6 seed, and the #4 seed plays the #5 seed in each conference. Since the seedings are based on win/loss, we found it appropriate to seed the teams based on Elo ratings (which is an outcome of win/loss). This also helps to avoid the complicated process should two teams have the same win loss record (there is a significantly lower probability two teams in this case have the exact same Elo rating). We simulated each matchup of the Playoffs to determine which teams advanced to ultimately identify the NBA Champions for the 2019-20 NBA season.

Predicting the 2019-20 NBA Season Award Winners

Another interesting part of any NBA season is the awards given to the players and teams based on their regular season performances. Some key awards that catch headlines every year include Most Valuable Player (MVP) and Defensive Player of the Year (DPOY). In addition to predicting the end-of-season standings, we decided that any analysis of the remainder of the 2019-20 NBA season would be incomplete unless it

discussed award winners in the major categories. As avid basketball fans, this idea resonated with us so we decided to scientifically predict who would ultimately win the MVP and DPOY awards.

In order to predict award winners at the end of the season, we needed to predict the leaders of some of the crucial statistical categories at the end of the season. To predict these players, we analyzed some significant statistics and identified the optimal regression model to predict these statistics. Some of the statistics we utilized in building this model include true shooting percentage (TS%), total rebound percentage (TRB%), assist percentage (AST%), and block percentage (BLK%) among 22 total recorded categories. For additional information on the exact statistics that were used in these predictions, refer to Table 3 (Appendix).

Accordingly, we examined different types of regression models in order to identify which type of model best predicted some of the major statistics. These include win shares (WS), value over replacement player (VORP), player efficiency rating (PER), usage percentage (USG%), offensive box plus/minus (OBPM), and defensive box plus/minus (DBPM). For more detailed explanations of the significance of each of these statistics, refer to Table 3 (Appendix).

To identify the best prediction model, we first predicted WS from the current 2019-20 season statistics using 4 different regression models: linear, lasso, ridge, and logistic. We measured the performance of every model with the actual WS values for each player using RMSE (root mean squared error) to determine which had the least error where smaller RMSE values indicated higher accuracy. In order to reduce Monte Carlo variability, we used 200 repeated random samples of the data for each model to find the true RMSE values.

We then used this to predict the MVP and DPOY by looking at the leaders at the end of the season in WS, VORP, PER, USG%, OBPM, and DBPM because these categories carried significant weighting in determining the respective awards. In order to produce statistics that would reflect the end-of-season data, we updated each player's stats based on their team, position, schedule matchups, and usage percentage. We weighted each of these features by category and used the current player statistics to simulate the expected statistics for every player at the end of the 2019-20 season. We chose these specific features because a player's team and schedule can heavily influence their output, the position they play directly affects which stats are affected, and their usage percentage indicates how heavily their team relies on that specific player (which correlates to more playing time). For example, a point guard is more likely to focus on assists, a shooting guard is more likely to focus on shooting percentage and 3-point attempt percentage, and a center is more likely to focus on rebounds and blocks. Similarly, a player with a high usage percentage will be instrumental to the team and will thus receive more playing time to add to their stat lines.

Another important note to consider is that some of the statistical categories are related to other categories. For example, offensive win shares (OWS) and defensive win shares (DWS) are directly related to overall win shares (WS) because they are simply more specific aspects of the general WS category. In order to account for these confounding variables and ensure that the predictions were accurately estimated from all of the relevant data, we made sure to exclude the respective confounding variables when running each regression model. For instance, we excluded BPM when running regression models on OBPM and DBPM for the same reason.

Results

First, we calculated the Elo ratings for each team for the games played so far in the 2019-20 season. As explained earlier, we incorporated 25% of the previous season's Elo ratings with 75% of this season's current Elo ratings. To better visualize how each team's Elo rating is updated throughout the season, Figure 2 shows the Elo rating of every team since the opening of the 2019-20 season for all games that have been played (prior to the suspension). This plot shows how the Elo rating can fluctuate over the course of a season and how the better teams tend to have an upward trajectory to build momentum towards clinching a playoff berth. We included a plot faceted by Conference because it better highlights which teams are likely to secure playoff seeding since the Playoffs matchups are organized by Conference.

Table 1: Elo Ratings for Every NBA Team (Descending)

Team	Elo Rating		Team	Elo Rating
Milwaukee Bucks	1628.4919	16	Sacramento Kings	1496.7678
Los Angeles Lakers	1627.8301	17	Brooklyn Nets	1491.0335
Toronto Raptors	1601.2836	18	Orlando Magic	1490.3970
Los Angeles Clippers	1588.4583	19	Portland Trail Blazers	1478.5000
Oklahoma City Thunder	1582.6036	20	San Antonio Spurs	1478.2600
Boston Celtics	1569.6068	21	Phoenix Suns	1453.5151
Denver Nuggets	1560.4563	22	Washington Wizards	1446.2512
Utah Jazz	1559.3314	23	Charlotte Hornets	1435.6679
Houston Rockets	1546.1499	24	New York Knicks	1430.9475
Indiana Pacers	1544.9614	25	Atlanta Hawks	1424.2177
Philadelphia 76ers	1536.7509	26	Cleveland Cavaliers	1412.5029
Miami Heat	1531.0399	27	Chicago Bulls	1410.1883
Dallas Mavericks	1530.4395	28	Minnesota Timberwolves	1391.4969
Memphis Grizzlies	1512.1054	29	Golden State Warriors	1387.3738
New Orleans Pelicans	1507.3374	30	Detroit Pistons	1383.5339

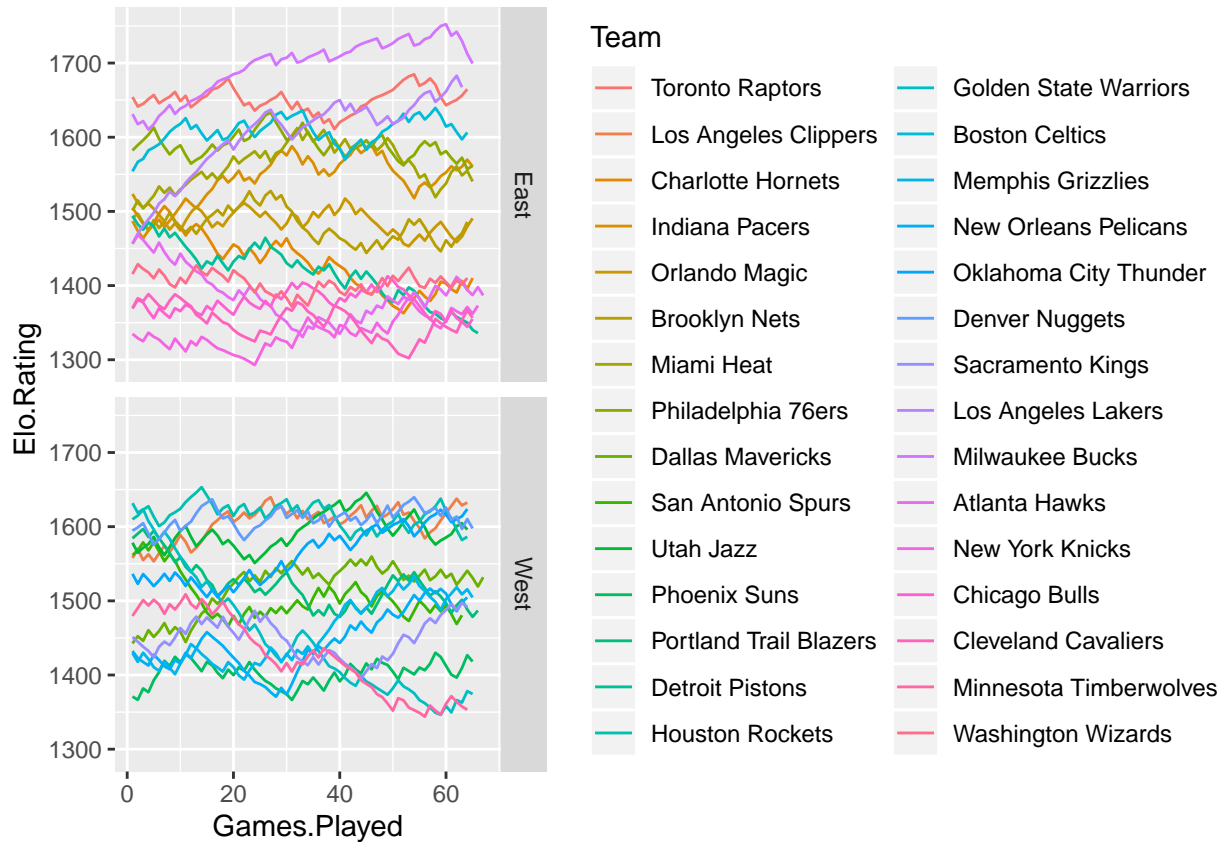


Figure 2: Elo Rating Over 2019-20 for Each NBA Team by Conference

We then predicted how the remaining of the regular season would play out and the standings of the teams at the end of the season. Table 1 shows the Elo ratings we predicted for each team at the end of the season. Noticeably, the teams who had already secured Playoff berths before the season suspension have higher Elo

rating predictions (as is to be expected).

Using the method explained above to predict the Playoff matchups, we were able to simulate the games for the Playoffs. This is how the first-round playoff matchups turned out:

Eastern Conference: Milwaukee vs. Orlando, Toronto vs. Brooklyn, Boston vs. Philadelphia, Miami vs. Indiana

Western Conference: LA Lakers vs. Memphis, LA Clippers vs. Dallas, Denver vs. Houston, Utah vs. Oklahoma

The same process for the regular season was applied to these matchups with a couple of tweaks. This simulation was a best-of-7 tournament (as are all playoffs and finals matches). As a result, we had to simulate the games until a team reached 4 wins. The teams who won and advanced to the second round are as follows:

Eastern Conference: Milwaukee, Toronto, Philadelphia, Miami

Western Conference: LA Lakers, LA Clippers, Houston, Utah

The second round is the same process as the first round, except this time, the top 4 teams of each conference advance, meaning that the other 4 teams are eliminated. The second round is also known as the Conference Semifinals, that is, the winners of these games advance to their respective Conference Finals to compete for the Conference title. The second round matchups are shown below:

Eastern Conference: Milwaukee vs. Miami, Toronto vs. Philadelphia

Western Conference: LA Lakers vs. Utah, LA Clippers vs. Houston

with the winners being:

Eastern Conference: Milwaukee, Toronto

Western Conference: LA Lakers, LA Clippers

The last round of the playoffs is the third round, also referred to as the Conference Finals. The top two teams from both the Western and Eastern Conference compete for the top spot in their respective Conference. The matchup for the Conference Finals was predicted to feature the heavy weight teams from both Conferences with the following matchups:

Eastern Conference: Milwaukee vs. Toronto

Western Conference: LA Lakers vs. LA Clippers

Again, these teams were put through a best-of-7 simulation, in which the following winners emerged:

Eastern Conference: Milwaukee

Western Conference: LA Lakers

Now that we had predicted our initial 16 Playoff teams down to the Conference Champions, we set the stage for the NBA Finals matchup between the Milwaukee Bucks and the Los Angeles Lakers. Both teams featured several superstars and were favored to meet in the NBA Finals by several sports analysts. **Simulating this exciting matchup, our simulation predicted that the NBA Championship would be won by the Los Angeles Lakers.**

Figure 3 shows how the Elo ratings for each NBA team changed throughout the end of the regular season as well as both the Playoffs and Finals. A line that remains flat for a short period of time represents the Elo rating remaining constant due to the fact of not playing at that instance. The changes to Elo rating come after they have played a game.

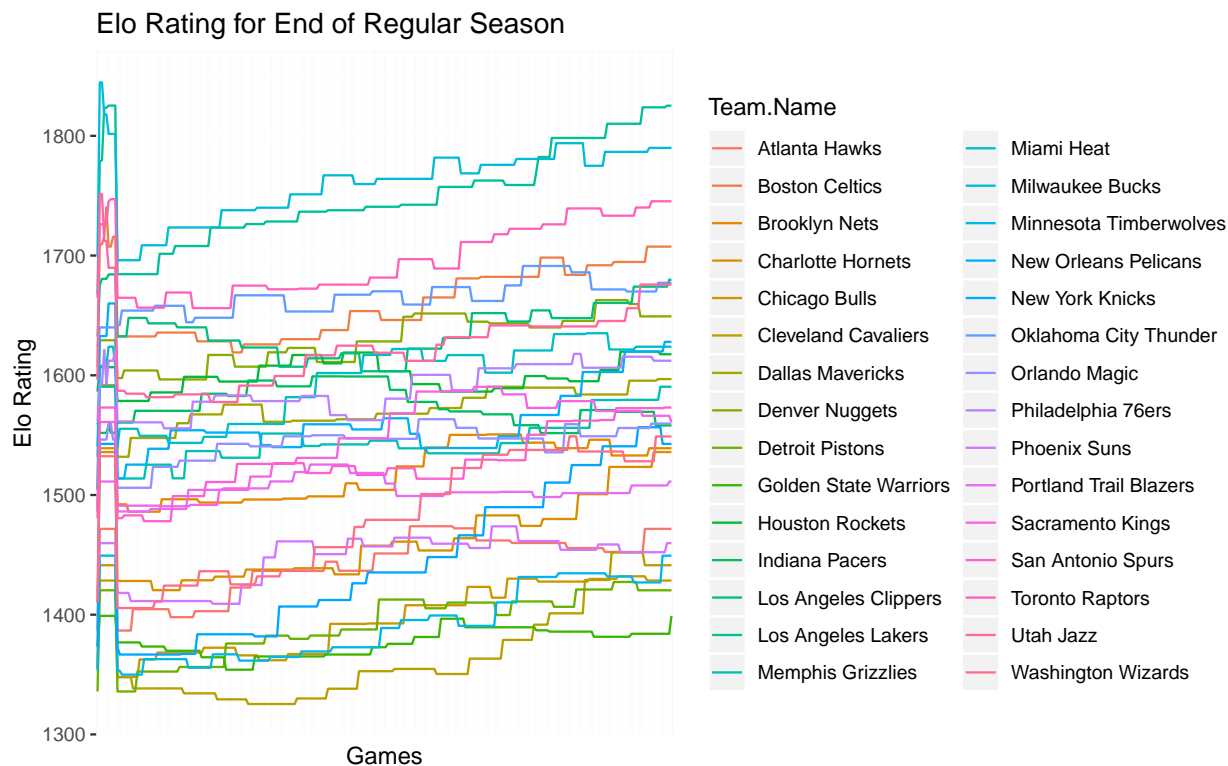


Figure 3: Elo Rating for End of Regular Season and Playoffs

In addition to simulating how the 2019-20 NBA season and Playoffs would have ended, we tested several regression models to determine the best model to predict end-of-season WS. By testing linear, lasso, ridge, and logistic regression models, found that the lasso model provided the lowest RMSE value (0.26119973) with the linear model (0.34557806) and logistic models (0.33962513) providing slightly higher RMSE values.

We then predicted the season leaders for the various principal categories of WS, VORP, PER, OBPM, and DBPM. We utilized a logistic regression model for predicting the stat leaders due to its relative simplicity and clarity. We verified that this logistic regression model was appropriate by measuring its RMSE values and plotting predicted end-of-season statistics for each player for every category as well as the actual vs. predicted statistics for each category for the current 2019-20 season data (Figure 4). The plots in Figure 4 indicate the end-of-season statistic value for each category based on a player's current 2019-20 season data for that statistic. The red lines indicate the value that a player can expect to end the season with for every value along the x -axis (which represents the player's current statistic for that category). These plots verify that the predicted stats fall within a very small margin of the actual stats and thus have small residual values (since a majority of this season's games have already been played). Also, there is a very minor shift in some of the data points (i.e. VORP) because the last month of games would likely change some of these player

Table 2: End-of-Season 2019-20 Predicted Stat Leaders with Logistic Regression Model

Player	Predicted WS	Player	Predicted VORP
Giannis Antetokounmpo	13.908386	Giannis Antetokounmpo	6.1050918
James Harden	13.133120	James Harden	5.4004209
Anthony Davis	11.389963	LeBron James	5.1503702
LeBron James	11.174695	Luka Dončić	4.8913897
Damian Lillard	10.916174	Kawhi Leonard	4.7491376
Player	Predicted PER	Player	Predicted USG%
Giannis Antetokounmpo	31.079413	James Harden	37.648219
James Harden	27.635517	Damian Lillard	36.649541
Luka Dončić	27.612389	Giannis Antetokounmpo	36.384652
Anthony Davis	27.256994	Luka Dončić	35.815117
Kawhi Leonard	27.238995	Bradley Beal	34.737061
Player	Predicted OBPM	Player	Predicted DBPM
James Harden	8.9420377	Giannis Antetokounmpo	3.7238770
Damian Lillard	8.6484498	Anthony Davis	3.1232859
Giannis Antetokounmpo	7.7182711	Kris Dunn	3.0098854
Luka Dončić	7.3270922	Nikola Jokić	2.6669762
LeBron James	7.1820940	Brook Lopez	2.6330744

statistics since teams with guaranteed playoff seeding are more likely to rest their star players.

Using these accumulated season statistics, we determined the end-of-season statistics by updating each players' stat lines based on team, position, schedule matchups, and usage percentage stats as explained in detail in the Methods section. The end-of-season stat leaders for each of the main categories is shown in Table 2. We verified these predictions by limiting the qualifiers for each category and comparing with each player's previous performance history. We narrowed the pool of players down by only considering players with more than 25 games played and 1200 minutes played because this reflects the criteria that the NBA Season Awards use to nominate qualifying players. By inspecting the top five players in each category, we noticed several household names and early season favorites for MVP and DPOY including Giannis Antetokounmpo, James Harden, LeBron James, and Anthony Davis.

In regards to determining the MVP nominees, the categories of WS, VORP, PER, and OBPM are significant because they constitute the most holistic perspective of each player's performance and are oriented around the 3 key factors for MVP: overall team success, value of the player to the team, and on-court presence. These 3 central components are acutely summarized in these statistics, so we used a point-based system to determine the top 3 nominees for MVP by looking at each player's rank within each category. In regards to determining the DPOY nominees, the categories of DBPM, USG%, PER, and VORP are significant because they constitute a comprehensive view of a player's defensive contributions. A DPOY nominee is typically a leader or near the top of several of these statistics, so these categories are a good representation of who is likely to be nominated for DPOY.

Giannis Antetokounmpo led an overwhelming majority of the statistics (WS, VORP, PER, USG%, DBPM). Not only did he lead each of these categories in our end-of-season predictions, he was also in the top three players for the other statistics, including OBPM shown in Table 2. **Due to this dominant stat line, we reasoned that our model predicted that Giannis Antetokounmpo would win both the MVP and DPOY awards.** This is an eye-catching prediction since this feat has only been accomplished twice before in the history of the NBA. From these projections, the reasonable runner-up for MVP was James Harden. For DPOY, these results show that the clear runner-up was Anthony Davis.

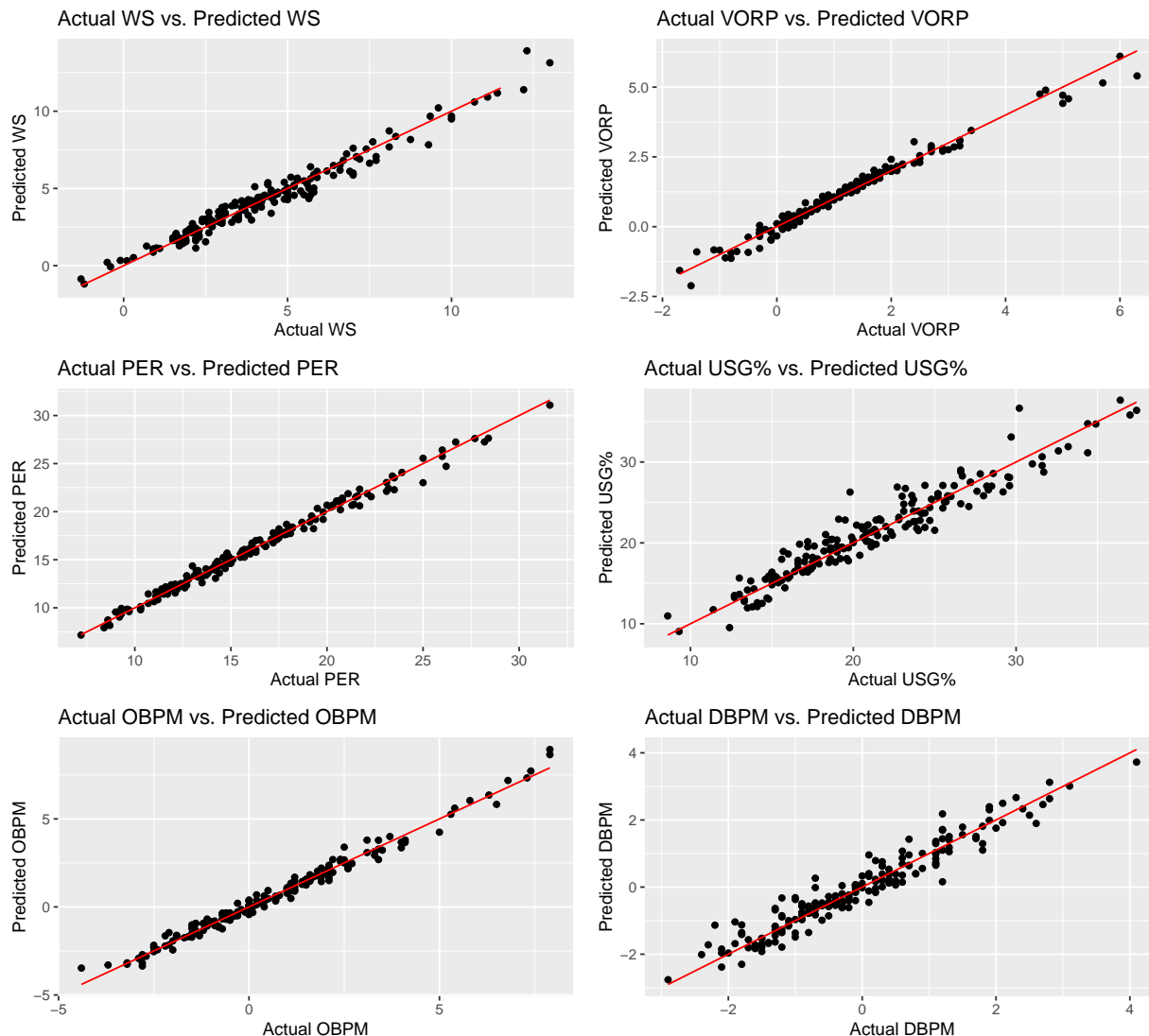


Figure 4: Predictions of Key Statistics with Logistic Regression Model

Conclusion

Through our analysis of the 2019-20 NBA season, we were able to analyze and predict several important aspects of the season had it continued as per schedule. We were first able to calculate the Elo rating for each team through the games that had already been played in the 2019-20 season. This was insightful because it showed us how each team performed over the course of the season. This would be interesting to GMs and sports analysts alike because it shows how a team can increase performance as the season goes on, remains at stagnant production (wins and losses about equal), or whether the team was affected by a major, unexpected hurdle in the middle of a season (i.e. injuries to role players, trades, etc.).

Using this understanding, we predicted how the season would end by simulating the remaining games and producing Elo ratings for the end of the season. This would be useful for a team manager or coach to evaluate the team's overall season performance in order to compare to previous seasons data and determine whether any changes need to be implemented, such as trading for certain positional players or releasing players that were ineffective.

In a similar manner, we predicted how the postseason Playoffs would shake out in terms of seeding, matchups, and winners. We ran another comparable prediction model using the end-of-season predicted

Elo ratings for each team. We ran these matchups in best-of-7 scenarios using the Elo probability factors explained above to identify which teams advanced. After running our predictions over numerous iterations, we predicted that the Western Conference Finals would feature a matchup between the LA Lakers and the LA Clippers and the Eastern Conference Finals would feature a matchup between the Milwaukee Bucks and the Toronto Raptors. Ultimately, the LA Lakers and Milwaukee Bucks emerged victorious for a Finals clash with the LA Lakers being crowned the predicted NBA Champions for 2019-20.

Not only did we predict how the season would end, we also used regression prediction models (linear, lasso, ridge, and logistic) to predict the end-of-season stat leaders for various crucial categories that are used to determine the MVP and DPOY. We found that the lasso regression model produced the lowest RMSE values when predicting WS based on the current 2019-20 data. Since the logistic regression model is easier to interpret and produced only a slightly higher RMSE value, we used the logistic regression model to predict the remaining key statistics (VORP, PER, USG%, OBPM, DBPM) for the end of the season. The projected leaders at the end of the season for each of these categories is depicted in Table 2. From these predictions, we found that Giannis Antetokounmpo is projected to win both the MVP and the DPOY with James Harden and Anthony Davis coming in second for each category, respectively. This is an noteworthy prediction because winning two of the most prestigious awards in basketball is an amazing testament to the player's performance this season. This information would be useful to GMs, franchise front offices, and the players' agents. For example, the franchise front offices could use this type of prediction to help build the best contract for the star players that would also benefit the team. Additionally, this would be useful to the players' agents because it would help them argue for favorable contract negotiations for their clients.

Overall, these types of predictions of season outcomes, Playoff matchups, and season award winners could immensely help teams with identifying how to improve for upcoming seasons or how to better prepare their players for one-on-one matchups against certain teams. Not only did this project yield entertaining and interesting results for the 2019-20 season, it could be further extended to future seasons when teams are trying to predict how to negotiate contracts for certain players.

Drawbacks of the Elo Rating Predictions

Although Elo ratings are simple and intuitive, this is both a blessing and a curse. Given the performance, Elo rating should be commended, however, this does not mean it should be used to support addictive sports gambling habits because it simply uses previous performance to predict future performance. This method is designed to be effective despite its simplicity. This model does not account for roster changes, injury updates, or individual player performance as most sports prediction models should. It evaluates the teams as a whole, solely based on win/loss, margin of victory, and home court advantage.

Also, the simulations are random, based on a team's probability of winning. This means the results could vary each time you run it, which requires running it a reasonable amount of times (at least 100 iterations) in order to get meaningful results. Since the predictions tend to vary a little with each run, after running it for multiple iterations, the consistent winner of the 2019-20 season proved to be the Los Angeles Lakers. This can be done by wrapping the script in a loop that runs it for a large number of iterations while keeping track of the winners of each season, and counting each team's wins to find who won most often. This method was used to get the result for the Los Angeles Lakers, but is not shown within this report due to its scope.

Appendix

ELO	EQUIVALENT RECORD	TEAM DESCRIPTION
1800	67-15	All-time great
1700	60-22	Title contender
1600	51-31	Playoff bound
1500	41-41	Average
1400	31-51	In the lottery
1300	22-60	LOL
1200	15-67	Historically awful

Figure 5: Breakdown of Elo Rating

Table 3: Statistics Recorded by NBA for Every Player

Statistics	Meaning
G	Num. games played
MP	Num. minutes played
PER	Measure of per-minute production
TS.	Overall shooting efficiency
X3PAr	% of field goal attempts from 3-point range
FTr	Num. free throw attempts per field goal attempt
ORB.	% of available offensive rebounds that a player grabbed
DRB.	% of available defensive rebounds that a player grabbed
TRB.	% of available total rebounds that a player grabbed
AST.	% of teammate field goals that a player assisted
STL.	% of opponent possessions that were stolen by a player
BLK.	% of opponent field goals attempts that were blocked by a player
TOV.	Num. turnovers committed per 100 plays
USG.	% of team plays used by a player
OWS	Num. wins contributed by a player from his offense
DWS	Num. wins contributed by a player from his defense
WS	Num. wins contributed by a player
WS.48	Num. wins contributed by a player per 48 minutes
OBPM	Offensive points per 100 possessions above a league-average player
DBPM	Defensive points per 100 possessions above a league-average player
BPM	Total points per 100 possessions above a league-average player
VORP	Points per 100 team possessions contributed by a player above a replacement-level player

References

- [1] List of all sporting events canceled around the world during the coronavirus pandemic (https://www.espn.com/olympics/story/_/id/28824781/list-sporting-events-canceled-coronavirus)
- [2] Elo ratings system (https://en.wikipedia.org/wiki/Elo_rating_system)
- [3] FiveThirtyEight NBA Elo Ratings (<https://fivethirtyeight.com/features/how-we-calculate-nba-elo-ratings/>)
- [4] Compilation of in-depth NBA statistics (<https://www.basketball-reference.com/>)
- [5] Elo Ratings for NBA Teams (<http://practicallypredictable.com/2018/04/15/elo-ratings-for-nba-teams/#more-1019>)