Finding the True MVP*

An NBA Analysis Through Linear Regression

Yan Mezhiborsky

December 3, 2024

This paper examines the methodology behind the NBA's Most Valuable Player (MVP) award, introducing the MVP Index as a data-driven measure to evaluate player performance. The MVP Index combines key statistics, including points, assists, rebounds, and win shares, to provide a comprehensive metric of value. Using historical data and statistical modeling, we identify discrepancies between actual MVP winners and those with the highest MVP Index, highlighting the influence of narrative and qualitative factors on voting outcomes. The findings underscore the need for greater transparency and consistency in the MVP selection process while offering a framework for more objective evaluations.

Table of contents

*Code and data are available at: [https://github.com/Mezhi18/NBAMVP.]

1 Introduction

Basketball's Most Valuable Player (MVP) award represents one of the most prestigious accolades in professional sports, honoring the individual deemed to have contributed the most to their team's success over a season. While the selection process has traditionally relied on expert voting and narrative-driven narratives, the question of whether the MVP truly reflects a player's value remains open to debate. This paper seeks to address this issue by evaluating the award through a data-driven lens, introducing the MVP Index as a novel measure of player impact. By exploring historical trends, inconsistencies, and predictive modeling, this research aims to shed light on how MVP selections align—or fail to align—with statistical measures of excellence.

At the heart of this analysis is the estimand, the MVP Index, which combines key player performance metrics into a single, normalized statistic. This index incorporates points, assists, rebounds, blocks, steals, and win shares, weighted to reflect their relative importance. The MVP Index provides a quantitative foundation for evaluating player performance, offering an alternative to the sometimes subjective criteria that influence MVP voting. The study employs statistical modeling, with linear regression used to predict the MVP Index for future seasons, enabling a comparison between the players with the best statistical performance and those who ultimately win the award.

The results of the analysis reveal several significant insights. In some seasons, the MVP Index aligns closely with the actual MVP winner, validating the selection process. However, there are also notable discrepancies, where players with superior statistical performances were overlooked in favor of candidates with compelling narratives or team success factors. These findings suggest that while data can provide clarity, the MVP award is often shaped by qualitative elements that extend beyond pure statistics.

Understanding these patterns is not only critical for fans and analysts but also holds implications for broader debates about transparency and fairness in sports. The MVP award significantly impacts player legacies, influencing Hall of Fame considerations, sponsorship deals, and even contract negotiations. A more robust and transparent framework for assessing MVP candidates could enhance the credibility of the award and ensure that it reflects the contributions of players who truly make a difference.

The remainder of this paper is structured as follows. Section 2 outlines the data sources, variables and processing steps that were use to help construct our very own `MVP Index`. Next, Section 3 talks about our model and how it was developped as well as how we are using it to predict the future statistics. Then, Section 4, we analyze MVP statistics over that nearly 40 years and how they have changed over time and highlights important discrepancies between our model and past MVPs. Lastly, Section 5 discusses the broader implications of these findings and acknowledges the paper's limitations.

2 Data

2.1 Overview

We use the statistical programming language R (R Core Team 2023), as well as the following packages to help clean and manipulate our data, (Wickham et al. 2023), (Goodrich et al. 2022), (Wickham, François, et al. 2023), (Firke 2023), (Grolemund and Wickham 2023), (Arel-Bundock 2023), (Xie 2023), (Carpenter et al. 2023), (Zhu 2023). Using these packages we cleaned our data, added years for the data sets that only had seasons, referring to the basketball season, as well as helped create some nice tables, unlike the traditional R software can create.

Our data was collected from (Sports Reference LLC 2024), we collected data from every player that received a vote for any All-NBA teams since 1986, this includes a vote cast for the First All-NBA Team, the Second All-NBA Team, and the Third All-NBA Team. As we believed that taking the players that received votes for the MVP awards would result in a smaller sample. We also created a new statistic call `MVP_Index` that will be defined in Section 2.2. In addition, we collected a data set for all NBA MVPs since 1986, so we could compare to the ones we have chosen with the `MVP_Index`.

Following Alexander (2023), we use techniques and methods from this text we have analyzed the data using models, graphs, tables and other methods of data analysis to conduct our research to find players that could have been MVPs and what the MVP for the following years will look like(statistically).

2.2 Variables

2.2.1 All NBA Variables

As we collected data from each year since 1986 individually, we have utilized the following variables:

- `Team`: Whether the player was in the 1st, 2nd, or 3rd All NBA Teams or OVR if they did not make the top three teams but still received a vote.
- `Pos`: The player's position including the generalized positions G, F, C, for Guard, Forward, and Center respectively.
- `Player`: The name of the NBA player.
- `Age`: The age of the NBA player.
- `Tm`: The three letter designation for the NBA team the player played for the year.
- `G`: Games played by the player.
- `MP`: Minutes played per game.
- `PTS`: Points per game.
- `TRB`: Total Rebounds per game, including offensive and defensive rebounds.
- `BLK`: Blocks per game
- `STL`: Steals Per game
- `WS`: The amount of wins allocated to the individual player

As each year has an individual data set we took PTS, TRB, AST, BLK, STL,a nd WS took the average for that year and normalized each player's statistics by dividing them by the average of that year with this we have created the MVP Index which we have defined as:

$$MVP_{Index} = 0.8PTS_{Norm} + 0.25AST_{Norm} + 0.25TRB_{Norm} + 0.1STL_{Norm} + 0.1BLK_{Norm} + 0.15WS_{Norm}$$

With this `MVP Index` we have created our own list of who could have been MVP base on statistics for that year, essentially comparing how good the player was to othe rplayers in the basketball season.

2.2.2 Historical MVP Variable

For our second set of date we have all MVPs since 1986 also sourced from (Sports Reference LLC 2024), and the variables that have been used are the following:

- `Player`: The name of the NBA player.
- `Year`: The year in which the player was selected as MVP.
- `Tm`: The three letter designation for the NBA team the player played for the year.
- `G`: Games played by the player.
- `MP`: Minutes played per game.
- `PTS`: Points per game.
- `TRB`: Total Rebounds per game, including offensive and defensive rebounds.
- `BLK`: Blocks per game
- `STL`: Steals Per game
- `WS`: The amount of wins allocated to the individual player

2.3 Measurement

Data collected in sports and more specifically basketball is rather simple. When an NBA game is being played there are people who's entire job is to keep track of a players statistics throughout the game, whenever a player makes a shot, or steals they ball it is being logged by that specific person. After the standard 82 games, there could be morte or less under different circumstances such ast COVID-19, injuries, or even trades all regular season statistics have been collected. As most of our in game statistics are based on the players 'per game' average, the total number of points blocks, etc., are recorded it is divided by the number of games the player played that season.

2.4 Outcome variable: MVP Index

Our outcome variable is the same as our estimand in Section 1. As `MVP Index` is dependent on our other statistics, Points, Assists, Rebounds, Steals, Blocks, and Win Shares. As we are studying MVPs in our paper this is our Variable of interest ast it represents how good a player was in comparison to the other best players in the NBA that year, from the sample of players that have received any ALL NBA Team vote.

3 Model

The goal of our modelling strategy is twofold. Firstly, we will be using a linear regression model to predict the MVP Index of the mvp in the following year regardless of the player. This will tell us how much better we can expect the 2025 MVP to be compared to his fellow

NBA players. Secondly, We will use another data set we have created by selecting players that had the highest MVP index in their respective year whether they had won MVP or not. So, Using both these predictions we will be able to compare whether we believe the player with the highest MVP index will in fact be the next NBA MVP.

Here we briefly describe the Linear analysis model used to investigate what the MVP Index of the next NBA MVP will be. For our model we will use collected data describe in Section 2.

Background details and diagnostics are included in Appendix D.

3.1 Model set-up

Define y_i as the MVP Index. Then π_i as the points per game, then α_i as assists per game, next we have ρ_i as rebounds per game. Furthermore, we have β_i as blocks per game, then we have ξ_i for steals per game, lastly we have ω_i for win shares. With these variables we will be predicting the MVP Index for the following NBA season.

$$
\begin{aligned}
y_i|\mu_i, \sigma &\sim \text{Normal}(\mu_i, \sigma) && (1)\\
\mu_i &= \pi_i + \alpha_i + \rho_i + \beta_i + \xi_i + \omega_i && (2)\\
\pi &\sim \text{Normal}(0, 2.5) && (3)\\
\alpha &\sim \text{Normal}(0, 2.5) && (4)\\
\rho &\sim \text{Normal}(0, 2.5) && (5)\\
\beta &\sim \text{Normal}(0, 2.5) && (6)\\
\xi &\sim \text{Normal}(0, 2.5) && (7)\\
\omega &\sim \text{Normal}(0, 2.5) && (8)\\
\sigma &\sim \text{Exponential}(1) && (9)
\end{aligned}
$$

3.1.1 Model justification

We expect a positive relationship between Between MVP Index and points, assists, rebounds, block, steals and win shares, as as these are all variables that define the MVP Index. In addition we expect a positive relationship as all of these statitcs indicate a good player, the larger anyone of these stats stats are the larger the MVP Index will be for the respective player. Just because a player does not aquire many block or many asists does not mean a player will have a low MVP rating, this just mean they will have to compensate with their other stats. we can also see in Section D that we have a p-value of less than 2.2×10^{-16}, which is very low and gives us confidence the the model is accurate.

Table 1

3.1.2 Model Results

Below we can see Table 1 where we can see the MVPs from the last 10 years, including an additional unknown MVP with a projected MVP Index score. For this table we used both Historical Data with past MVP winners as well as data we have created on our own. The data we have created includes players that have had the highest MVP Index within their respective year, so not necessarily the same player that truly won the MVP Award that year. As we can see the historical data predicts a player to have an MVP index of 2.40 while our data suggests that the Player with the highest MVP index will be 2.45, as such it is very likely that the player with the highest MVP Index will not be the player selected for the MVP award. As with the 2024 season the player awarded with the MVP award was Nikola Jokic with an MVP Index score of 2.16, while Luka Doncic another player well deserving of the MVP Award and was in the race to win the award has a score of 2.20 which is in fact higher than Nikola Jokic's.

As we have learned from Table 1 and which we can see in **?@sec-index** it is obvious that the MVP is not always the player that performs the best statistically, as LeBron James would have won 8 MVP awards before entering the 2013-2014 season, which would be 8 MVP awards in 10 years. Therefore, there must other factors excluding the statistics we were looking at, such as human error or prejudice, team success, and overall story telling which we will look further into in Section 5.

4 Results

In this section we will be taking a look at some statistics since 1986, starting with MVP index, the moving on to PPG, the RPG, and lastly we will look at APG. In these graphs we will be looking at Historical MVPs as well as players from each year with the highest MVP index. Analyzing these graphs will help us will help us understand how these stats developped through time among the best of the best in the NBA and potentially which statistics are valued more in real life than in the real MVP race.

4.1 MVP Index Since 1986

In Figure 1 we see that both graphs seem to bounce around a lot until we reach the 2020s where both seem to converge at around an MVP Index score of 2.2. This could mean several things the MVPs are getting worse which is incredibly unlikely and we do not believe that this is the case, another possibility is that the competition is getting better and more consistent. We must remember that the MVP index relies just as much on all the other players in the

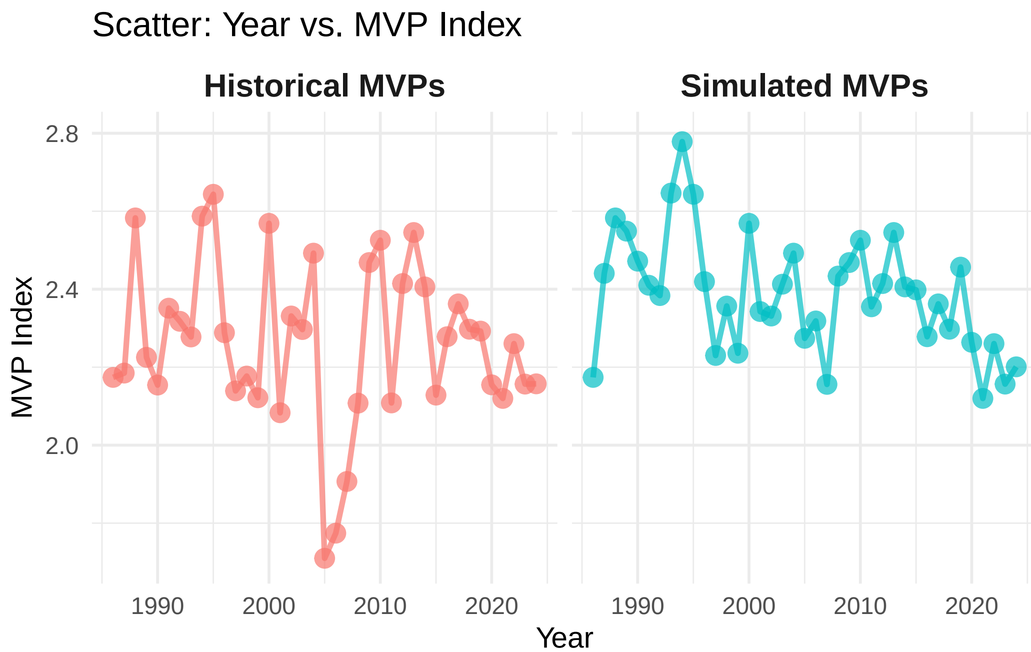

Figure 1: MVP index of of Historical and simulated MVPs

data set as much as it relies on the player who's MVP Index we are looking at. If we look at the 2004-05 and the 2005-06 seasons we can see a large dip in historical MVPs, this can be attributed to Phoenix Suns' point guard Steve Nash, as he was not even in the top 10 of MVP index with two teammate Amar'e Stoudemire and Shawn Marion two of Nash's teammates both ahve a higher Index than Nash. Although these are some of the most contentious MVP selections in the past 20 years, it must be considered what NBA analysts saw in him, as his team did have success, although falling short of the championship, and he was a great leader both on and off the court, as a floor general and a mentor. The question still remains did Nash deserve those MVPs? And is the Selection process for choosing an MVP flawed?

4.2 PPG since 1986

We will no be looking at Figure 2 where we are looking at MVP's points per game through the years. As this graph is similar to Figure 1 in the sense that then numbers do seem to be bouncing around although here we see that they do not seem to be converging around a single point. Although in this grapgh we see something interesting as well, Both Simulated MVs and Historical MVPs there seems to be an increasing baseline as well as an increasing top line, while the points are bouncing up and down through the years there along the top data points and the bottom data points, excluding Steve Nash, the only player to win MVP while scoring under 20ppg since 1986, who we discussed in Section 4.1. The interesting thing is that this occurs in both graphs, this tells us that as the years go on the points scored by MVPs, even

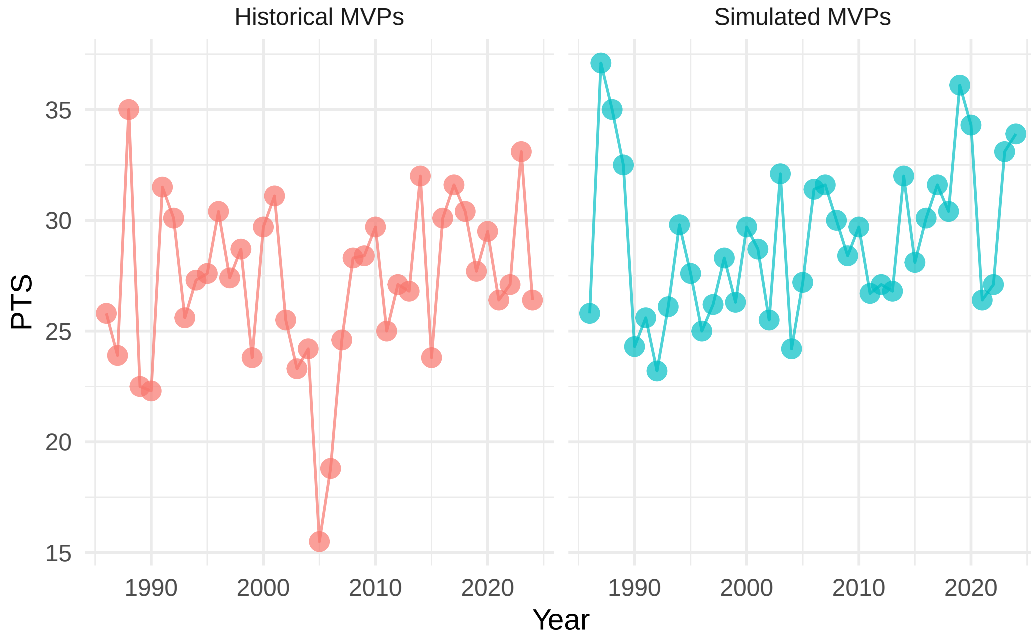

Figure 2: MVP PPG over Time

those who tend to score less are increasing this can be confirmed in a the paper (Mezhiborsky 2024).

4.3 RPG since 1986

In Figure 3 we are considering at rebounds per game. This gives us some additional information on the type of MVPs and what was prioritized at which points in time. We can see that there is a significant dip between the mid 2000s and the mid 2010s when we look at Historical Data and simulated MVPs there is a lack of 'Big Men', Centers and Power Forwards, as they are often the players with the most rebounds on a given team especially until the late 2010s, when smaller players such as Russell Westbrook Began recording more more rebounds. During this time there seems to be a vacuum of elite big men, from Shaquille O'neal, Tim Duncan, and Kevin Garnett in the early 2000s to the void we see above until the late 2010s and early 2020s where Giannis Antetekumpo, Nikola Jokic, and Joel Embiid rose to power and one of the three previously mentioned players won the MVP award in the last 6 years, which is why we can see the big jump in rebounds in the last few years of the graphs.

4.4 APG since 1986

Lastly we will be looking at Figure 4 in which we graph assists per game over time since 1986. Just as in Figure 3 we have a dip in assists per game among MVPs and Simulated MVPs in

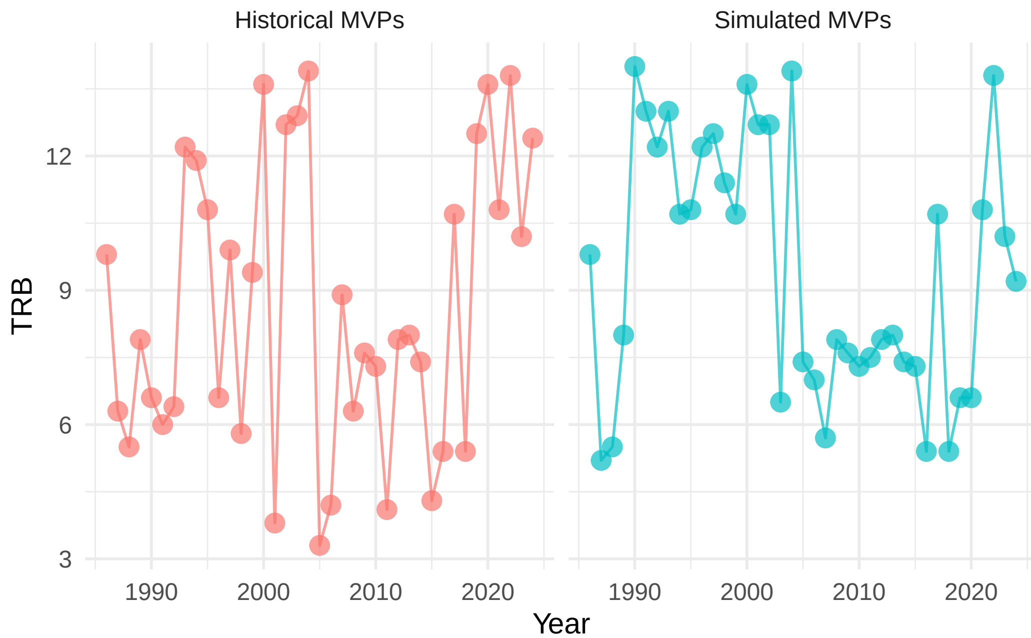

Figure 3: MVP RPG over Time

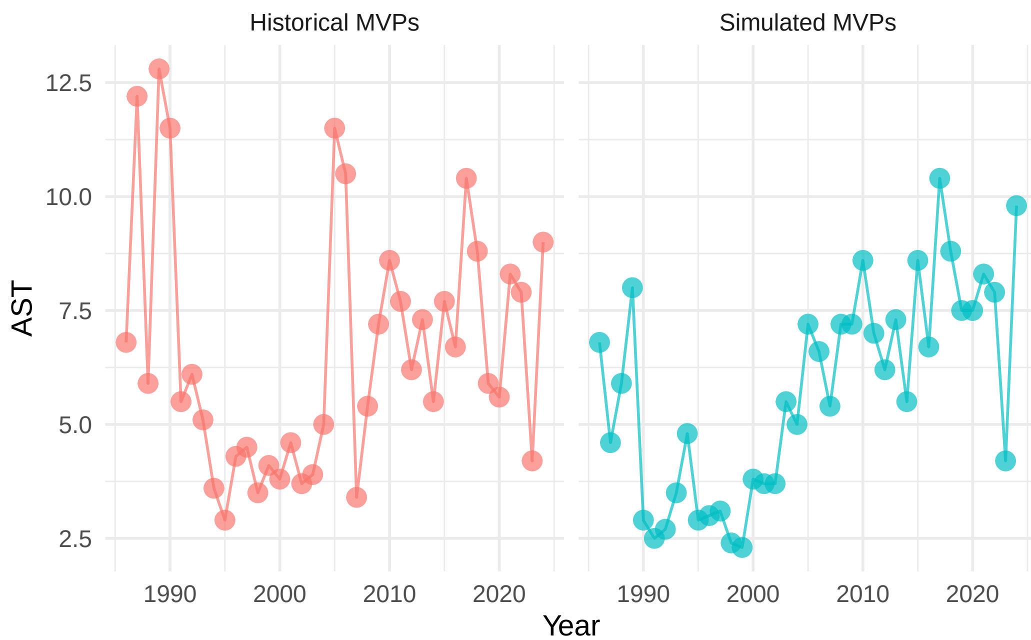

Figure 4: MVP ASG over Time

which the assists per game in the 1990s and the early 2000s, where big men, while know for their rebounds and not so much for their assists were dominating the league. While we see a gradual increase in the 2020s and 2020s with the small outlier of Joel Embiid in 2023. After looking at these last few visuals we can come to the conclusion that each of these statistics seem to have a somewhat of a growing trend whether consistent or not it does seem to exists, we can come to the conclusions that MVPs in the modern NBA are more likely to be selected if they are a well rounded player, a player that score a lot of points, rebound the ball, and be an unselfish player by by passing the ball a lot especially in scoring situations.

5 Discussion

5.1 Human Error and Bias

We must remember that at the end of the day the MVP is selected by a group of NBA analysts that come together once a year and cast their ballot for who they think the MVP should be. We must also remember that the award is called the 'Most Valuable Player' Award and not the 'Best Player' Award indicating that the player that should be selected is more valuable to their team than any other player. As people are deciding the player to receive these awards, it is good as they are able to take into account quantifiable and intangibles that machines would not, such as leadership, teammate morale, on-court behavior, off-court behavior and many more attributes that could affect their vote, as a machine would likely create a and MVP Index not completely unlike ours and choose the highest or lowest number depending on how it is measured.On the other hand, humans are extremely prone to bias, for example Stephen Curry was named the first unanimous MVP in 2016 when he received all 131 votes cast that year, while in 2013 LeBron James received 120 out of 121 votes, the last vote went to Carmelo Anthony, who did not have a bad year by any means, but LeBron's Greatness that year was undeniable. Lebron's incredible statistical performance that year was one thing to behold, but he would go on to win his second NBA championship against one the most put together teams ever assembled. This shows that no matter how uncorrupt and trust worthy a set of analysts may seem they are human at the end of the day and are prone to error at the least and bias at the worst.

Sometimes choosing an mvp that would make a good story and in utn make the NBA more money also occasionally happens, for example when derrick rose won the MVP award in 2011 few people would make the statement that he was the best player in the world, as in almost every statistical category Lebron James was superior and Lebron's team, the Miami Heat, had more playoff success that year. An additional point, Derrick rose was the home town hero in Chicago, after growing in the windy city he was drafted by his hometown team, and attempted to regain the success the team had in the !990s with Michael Jordan. This makes an incredible story.

5.2 Consistency vs. Peak Performance

Averages do not show everything. It is possible for a player to be incredibly inconsistent for their team, playing well against weaker opponents while struggling against stronger ones. This could lead to impressive average stats that fail to reflect their performance when the team needed them most. On the other hand, peak performance, where a player dominates in key moments, often captures attention and can overshadow an otherwise inconsistent season. The MVP debate often centers around this tension between reliability and brilliance.A player who delivers steady, high-level contributions throughout the season provides a reliable foundation for their team's success. These players might not always have headline-grabbing moments, but their ability to perform well in nearly every game can be the difference between making the playoffs or falling short. Their contributions build trust among teammates and coaches, as they can be relied upon to deliver in critical situations time and time again.

In contrast, players who showcase moments of sheer brilliance often become the face of the league. These athletes might not exhibit the same game-to-game reliability, but their ability to dominate in spectacular fashion leaves a lasting impression on fans and voters alike. Peak performers can change the momentum of a season with game-winning shots, record-breaking performances, or memorable battles against top opponents. However, their inconsistency can leave teams vulnerable during less-publicized matchups or critical stretches of the season.Ultimately, the MVP conversation remains an evolving debate, shaped by both quantitative measures and subjective interpretations of what makes a player "most valuable."

5.3 Voting Methodology and Transparency

The voting methodology and transparency in the NBA, particularly for awards like the MVP, have significant implications for the league's credibility and fan engagement. The MVP award is determined by a panel of sportswriters and broadcasters from the United States and Canada, with each voter ranking their top five candidates. Points are assigned on a weighted scale, and the player with the highest total is named MVP. While this system allows for expert analysis, transparency remains a recurring discussion point. Critics often call for more public access to the detailed ballots to ensure accountability and minimize potential biases. The NBA has made strides in recent years by releasing voters' choices post-award, fostering a culture of openness. However, debates about the inclusion of diverse perspectives, potential regional biases, and the weight given to narrative versus statistical performance continue to shape the conversation around fairness and transparency in the voting process.

5.4 Weaknesses and next steps

First, what would have been ideal especially in terms of using our model, would have been to take players from the all 3 All NBA teams, and download the data from these players and have our model analyze the data so we would have an actual name of a player for our predicted

MVP Index and not just the score. While this would have been great time constraints and other external factors hindered us from implementing this further analysis.

Second, for a more accurate MVP Index we could have sampled the entire league instead of the top 20-30 players, as this would have made a more accurate model as well as made the great players look much more exceptional with regards to the MVP Index, as we have only sampled players who have received at least one vote for an All NBA team, this leads to the average statistics we used to normalize our stats. This ended up making the Phenomenal players look great, and the great players look average, and the good players look sub par.

Next, team records are also incredibly important for MVP, if a player played incredibly one season but barely made the playoffs, or even did not at all, the player likely should not be receiving the MVP award, as how much worse could said team get if they weren't that good with the so called "best" player in the league. Unfortunately we were notable to inclue team record, team win percentage, or even player win percentage if we do not want to include the games a player may have missed. Unfortunately this data was not included in either type of data set that we sourced from (Sports Reference LLC 2024), and due to time constraints downloading potentially dozens if not hundreds additional data sets did not seem to be the best use of time in our case.

Lastly, as mentioned in Section 5.1 the MVP award does not solely rely on statistics and quantifiable factors. While, stats are heavily used in the decision for choosing the MVP there are other. non tangible factors that must be taken into account, other wise incredibly selfish players who do not help their teams have a much higher chance of winning.

Our next steps, would like be to monitor this NBA season closely potentially add some of the data mentioned earlier in Section 5.4, such as win percentage, and individual historical statistics to more accurate predict the model. We would also like to revisit this paper tat the end of this NBA season once the NBA awards are announced and see how we could improve our model and how accurate it was in predicting the MVP Index of the next MVP.

Appendix

A Surveys, Sampling, and Observational Data

A.1 Overview of MVP Voting Methodology

The MVP award voting process relies on a panel of sportswriters and broadcasters across the United States and Canada, with each voter ranking their top five players. The point system allocates 10 points for a first-place vote, 7 for second, 5 for third, 3 for fourth, and 1 for fifth. The player with the highest total points is named the MVP. While this system ensures broad coverage across media markets, it introduces potential biases rooted in the selection of voters, the weighting system, and the lack of transparency in the criteria used by individual voters.

A.2 Observational Nature of NBA Data

MVP voting data is inherently observational, reflecting subjective decisions rather than controlled experiments. This reliance on subjective assessments introduces variability, as voters may prioritize narrative driven criteria such as team success, leadership, or memorable moments over statistical performance. Observational data like this cannot account for unmeasured factors, such as voters' implicit biases or media-driven narratives. Additionally, players on historically successful teams or those in larger markets may receive disproportionate attention, skewing the outcomes.

A.3 Sampling Representation of Voters

The panel of voters is not a random sample of all stakeholders in the NBA community. Instead, it is a curated group of media professionals, which may underrepresent certain perspectives, such as fans, coaches, or analytics experts. This introduces a potential sampling bias that could influence the results. For example, regional biases might favor players from large-market teams, while recency bias could elevate players with late-season surges. Simulation studies could explore the extent of these biases by randomly reweighting votes or introducing new voter segments, e.g., analytics specialists.

##Simulation of MVP Voting Under Different Scenarios

To examine the robustness of MVP outcomes, simulations were conducted to evaluate how changes to the voting methodology could influence results. For instance, adjusting the weight given to statistical performance in vote tallies revealed several seasons where the MVP Index-aligned winner differed from the historically selected MVP. Similarly, experimenting with alternative weighting schemes, such as assigning equal points for all top-five rankings instead of a tiered system, led to notable shifts in the final outcomes. These simulations highlight the significant role that methodological choices play in shaping MVP selections, showing that even small adjustments can result in different winners. By simulating different voting scenarios, it becomes clear that the current system is sensitive to how votes are counted and weighted, underscoring the need for transparency and careful consideration of methodology.

B Table By MVP Index

	Year	Player	MVP_index	PTS	AST	TRB	BLK	STL	WS	Source
1	1986	Larry Bird	2.173793	25.8	6.8	9.8	0.6	2.0	15.8	Simulated MVPs
2	1987	Michael Jordan	2.440668	37.1	4.6	5.2	1.5	2.9	16.9	Simulated MVPs
3	1988	Michael Jordan	2.582764	35.0	5.9	5.5	1.6	3.2	21.2	Simulated MVPs
4	1989	Michael Jordan	2.548866	32.5	8.0	8.0	0.8	2.9	19.8	Simulated MVPs
5	1990	Hakeem Olajuwon	2.471915	24.3	2.9	14.0	4.6	2.1	11.2	Simulated MVPs
6	1991	David Robinson	2.410097	25.6	2.5	13.0	3.9	1.5	17.0	Simulated MVPs
7	1992	David Robinson	2.383681	23.2	2.7	12.2	4.5	2.3	13.9	Simulated MVPs
8	1993	Hakeem Olajuwon	2.646521	26.1	3.5	13.0	4.2	1.8	15.8	Simulated MVPs
9	1994	David Robinson	2.778428	29.8	4.8	10.7	3.3	1.7	20.0	Simulated MVPs
10	1995	David Robinson	2.643401	27.6	2.9	10.8	3.2	1.7	17.5	Simulated MVPs
11	1996	David Robinson	2.419458	25.0	3.0	12.2	3.3	1.4	18.3	Simulated MVPs
12	1997	Shaquille O'Neal	2.229996	26.2	3.1	12.5	2.9	0.9	8.0	Simulated MVPs
13	1998	Shaquille O'Neal	2.356842	28.3	2.4	11.4	2.4	0.7	10.2	Simulated MVPs
14	1999	Shaquille O'Neal	2.235782	26.3	2.3	10.7	1.7	0.7	9.0	Simulated MVPs
15	2000	Shaquille O'Neal	2.568962	29.7	3.8	13.6	3.0	0.5	18.6	Simulated MVPs
16	2001	Shaquille O'Neal	2.342722	28.7	3.7	12.7	2.8	0.6	14.9	Simulated MVPs
17	2002	Tim Duncan	2.331213	25.5	3.7	12.7	2.5	0.7	17.8	Simulated MVPs
18	2003	Tracy McGrady	2.412863	32.1	5.5	6.5	0.8	1.7	16.1	Simulated MVPs
19	2004	Kevin Garnett	2.492411	24.2	5.0	13.9	2.2	1.5	18.3	Simulated MVPs
20	2005	LeBron James	2.274034	27.2	7.2	7.4	0.7	2.2	14.3	Simulated MVPs
21	2006	LeBron James	2.318547	31.4	6.6	7.0	0.8	1.6	16.3	Simulated MVPs
22	2007	Kobe Bryant	2.155932	31.6	5.4	5.7	0.5	1.4	13.0	Simulated MVPs
23	2008	LeBron James	2.433659	30.0	7.2	7.9	1.1	1.8	15.2	Simulated MVPs
24	2009	LeBron James	2.468354	28.4	7.2	7.6	1.1	1.7	20.3	Simulated MVPs
25	2010	LeBron James	2.525520	29.7	8.6	7.3	1.0	1.6	18.5	Simulated MVPs
26	2011	LeBron James	2.354880	26.7	7.0	7.5	0.6	1.6	15.6	Simulated MVPs
27	2012	LeBron James	2.414567	27.1	6.2	7.9	0.8	1.9	14.5	Simulated MVPs
28	2013	LeBron James	2.545409	26.8	7.3	8.0	0.9	1.7	19.3	Simulated MVPs
29	2014	Kevin Durant	2.405868	32.0	5.5	7.4	0.7	1.3	19.2	Simulated MVPs
30	2015	Russell Westbrook	2.398289	28.1	8.6	7.3	0.2	2.1	10.6	Simulated MVPs
31	2016	Stephen Curry	2.278181	30.1	6.7	5.4	0.2	2.1	17.9	Simulated MVPs
32	2017	Russell Westbrook	2.362328	31.6	10.4	10.7	0.4	1.6	13.1	Simulated MVPs
33	2018	James Harden	2.297349	30.4	8.8	5.4	0.7	1.8	15.4	Simulated MVPs

34	2019	James Harden	2.456584	36.1	7.5	6.6	0.7	2.0	15.2	Simulated MVPs
35	2020	James Harden	2.263751	34.3	7.5	6.6	0.9	1.8	13.1	Simulated MVPs
36	2021	Nikola Jokić	2.119867	26.4	8.3	10.8	0.7	1.3	15.6	Simulated MVPs
37	2022	Nikola Jokić	2.260379	27.1	7.9	13.8	0.9	1.5	15.2	Simulated MVPs
38	2023	Joel Embiid	2.156639	33.1	4.2	10.2	1.7	1.0	12.3	Simulated MVPs
39	2024	Luka Dončić	2.200904	33.9	9.8	9.2	0.5	1.4	12.0	Simulated MVPs

C Historical MVPs

	Season	Lg	Player	Voting	Age	Tm	G	MP	PTS	TRB	AST	STL
2	2023-24	NBA	Nikola Jokić	(V)	28	DEN	79	34.6	26.4	12.4	9.0	1.4
3	2022-23	NBA	Joel Embiid	(V)	28	PHI	66	34.6	33.1	10.2	4.2	1.0
4	2021-22	NBA	Nikola Jokić	(V)	26	DEN	74	33.5	27.1	13.8	7.9	1.5
5	2020-21	NBA	Nikola Jokić	(V)	25	DEN	72	34.6	26.4	10.8	8.3	1.3
6	2019-20	NBA	Giannis Antetokounmpo	(V)	25	MIL	63	30.4	29.5	13.6	5.6	1.0
7	2018-19	NBA	Giannis Antetokounmpo	(V)	24	MIL	72	32.8	27.7	12.5	5.9	1.3
8	2017-18	NBA	James Harden	(V)	28	HOU	72	35.4	30.4	5.4	8.8	1.8
9	2016-17	NBA	Russell Westbrook	(V)	28	OKC	81	34.6	31.6	10.7	10.4	1.6
10	2015-16	NBA	Stephen Curry	(V)	27	GSW	79	34.2	30.1	5.4	6.7	2.1
11	2014-15	NBA	Stephen Curry	(V)	26	GSW	80	32.7	23.8	4.3	7.7	2.0
12	2013-14	NBA	Kevin Durant	(V)	25	OKC	81	38.5	32.0	7.4	5.5	1.3
13	2012-13	NBA	LeBron James	(V)	28	MIA	76	37.9	26.8	8.0	7.3	1.7
14	Dec-11	NBA	LeBron James	(V)	27	MIA	62	37.5	27.1	7.9	6.2	1.9
15	Nov-10	NBA	Derrick Rose	(V)	22	CHI	81	37.4	25.0	4.1	7.7	1.0
16	Oct-09	NBA	LeBron James	(V)	25	CLE	76	39.0	29.7	7.3	8.6	1.6
17	Sep-08	NBA	LeBron James	(V)	24	CLE	81	37.7	28.4	7.6	7.2	1.7
18	Aug-07	NBA	Kobe Bryant	(V)	29	LAL	82	38.9	28.3	6.3	5.4	1.8
19	Jul-06	NBA	Dirk Nowitzki	(V)	28	DAL	78	36.2	24.6	8.9	3.4	0.7
20	Jun-05	NBA	Steve Nash	(V)	31	PHO	79	35.4	18.8	4.2	10.5	0.8
21	May-04	NBA	Steve Nash	(V)	30	PHO	75	34.3	15.5	3.3	11.5	1.0
22	Apr-03	NBA	Kevin Garnett	(V)	27	MIN	82	39.4	24.2	13.9	5.0	1.5
23	Mar-02	NBA	Tim Duncan	(V)	26	SAS	81	39.3	23.3	12.9	3.9	0.7
24	Feb-01	NBA	Tim Duncan	(V)	25	SAS	82	40.6	25.5	12.7	3.7	0.7
25	Jan-00	NBA	Allen Iverson	(V)	25	PHI	71	42.0	31.1	3.8	4.6	2.5
26	1999-00	NBA	Shaquille O'Neal	(V)	27	LAL	79	40.0	29.7	13.6	3.8	0.5
27	1998-99	NBA	Karl Malone	(V)	35	UTA	49	37.4	23.8	9.4	4.1	1.3
28	1997-98	NBA	Michael Jordan	(V)	34	CHI	82	38.8	28.7	5.8	3.5	1.7
29	1996-97	NBA	Karl Malone	(V)	33	UTA	82	36.6	27.4	9.9	4.5	1.4
30	1995-96	NBA	Michael Jordan	(V)	32	CHI	82	37.7	30.4	6.6	4.3	2.2
31	1994-95	NBA	David Robinson	(V)	29	SAS	81	38.0	27.6	10.8	2.9	1.7
32	1993-94	NBA	Hakeem Olajuwon	(V)	31	HOU	80	41.0	27.3	11.9	3.6	1.6

33	1992-93	NBA	Charles Barkley	(V)	29	PHO	76	37.6	25.6	12.2	5.1	1.6
34	1991-92	NBA	Michael Jordan	(V)	28	CHI	80	38.8	30.1	6.4	6.1	2.3
35	1990-91	NBA	Michael Jordan	(V)	27	CHI	82	37.0	31.5	6.0	5.5	2.7
36	1989-90	NBA	Magic Johnson	(V)	30	LAL	79	37.2	22.3	6.6	11.5	1.7
37	1988-89	NBA	Magic Johnson	(V)	29	LAL	77	37.5	22.5	7.9	12.8	1.8
38	1987-88	NBA	Michael Jordan	(V)	24	CHI	82	40.4	35.0	5.5	5.9	3.2
39	1986-87	NBA	Magic Johnson	(V)	27	LAL	80	36.3	23.9	6.3	12.2	1.7
40	1985-86	NBA	Larry Bird	(V)	29	BOS	82	38.0	25.8	9.8	6.8	2.0

	BLK	FG.	X3P.	FT.	WS	WS.48
2	0.9	0.583	0.359	0.817	17.0	0.299
3	1.7	0.548	0.330	0.857	12.3	0.259
4	0.9	0.583	0.337	0.810	15.2	0.296
5	0.7	0.566	0.388	0.868	15.6	0.301
6	1.0	0.553	0.304	0.633	11.1	0.279
7	1.5	0.578	0.256	0.729	14.4	0.292
8	0.7	0.449	0.367	0.858	15.4	0.289
9	0.4	0.425	0.343	0.845	13.1	0.224
10	0.2	0.504	0.454	0.908	17.9	0.318
11	0.2	0.487	0.443	0.914	15.7	0.288
12	0.7	0.503	0.391	0.873	19.2	0.295
13	0.9	0.565	0.406	0.753	19.3	0.322
14	0.8	0.531	0.362	0.771	14.5	0.298
15	0.6	0.445	0.332	0.858	13.1	0.208
16	1.0	0.503	0.333	0.767	18.5	0.299
17	1.1	0.489	0.344	0.780	20.3	0.318
18	0.5	0.459	0.361	0.840	13.8	0.208
19	0.8	0.502	0.416	0.904	16.3	0.278
20	0.2	0.512	0.439	0.921	12.4	0.212
21	0.1	0.502	0.431	0.887	10.9	0.203
22	2.2	0.499	0.256	0.791	18.3	0.272
23	2.9	0.513	0.273	0.710	16.5	0.248
24	2.5	0.508	0.100	0.799	17.8	0.257
25	0.3	0.420	0.320	0.814	11.8	0.190
26	3.0	0.574	0.000	0.524	18.6	0.283
27	0.6	0.493	0.000	0.788	9.6	0.252
28	0.5	0.465	0.238	0.784	15.8	0.238
29	0.6	0.550	0.000	0.755	16.7	0.268
30	0.5	0.495	0.427	0.834	20.4	0.317
31	3.2	0.530	0.300	0.774	17.5	0.273
32	3.7	0.528	0.421	0.716	14.3	0.210
33	1.0	0.520	0.305	0.765	14.4	0.242
34	0.9	0.519	0.270	0.832	17.7	0.274
35	1.0	0.539	0.312	0.851	20.3	0.321

```
36 0.4 0.480 0.384 0.890 16.5 0.270
37 0.3 0.509 0.314 0.911 16.1 0.267
38 1.6 0.535 0.132 0.841 21.2 0.308
39 0.5 0.522 0.205 0.848 15.9 0.263
40 0.6 0.496 0.423 0.896 15.8 0.244
```

D Model details

```
Call:
lm(formula = MVP_index ~ PTS + AST + TRB + BLK + STL + WS, data = nba_master)

Residuals:
     Min       1Q   Median       3Q      Max
-0.27967 -0.06232 -0.00013  0.06198  0.33044

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.1545172  0.0136329   11.33   <2e-16 ***
PTS         0.0337247  0.0005182   65.08   <2e-16 ***
AST         0.0445034  0.0012761   34.88   <2e-16 ***
TRB         0.0295472  0.0010966   26.95   <2e-16 ***
BLK         0.1125198  0.0041166   27.33   <2e-16 ***
STL         0.1060966  0.0056427   18.80   <2e-16 ***
WS          0.0199776  0.0009258   21.58   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09672 on 1536 degrees of freedom
Multiple R-squared:  0.8992,    Adjusted R-squared:  0.8988
F-statistic:  2284 on 6 and 1536 DF,  p-value: < 2.2e-16
```

References

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

Arel-Bundock, Vincent. 2023. *Modelsummary: Beautiful and Customizable Model Summaries and Tables*. https://CRAN.R-project.org/package=modelsummary.

Carpenter, Bob, Marco Inacio, Mitzi Morris, et al. 2023. *StanHeaders: C++ Header Files for Stan*. https://CRAN.R-project.org/package=StanHeaders.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "rstanarm: Bayesian applied regression modeling via Stan." https://mc-stan.org/rstanarm/.

Grolemund, Garrett, and Hadley Wickham. 2023. *Lubridate: Make Dealing with Dates a Little Easier*. https://CRAN.R-project.org/package=lubridate.

Mezhiborsky, Yan. 2024. "The Future of the NBA Is on the Horizon." https://github.com/Mezhi18/NBAExpansion.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Sports Reference LLC. 2024. Basketball-Reference.com - Basketball Statistics and History. https://www.basketball-reference.com/.

Wickham, Hadley et al. 2023. *The Tidyverse: Easily Install and Load the 'Tidyverse'*. https://CRAN.R-project.org/package=tidyverse.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://CRAN.R-project.org/package=knitr.

Zhu, Hao. 2023. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.