Estimating Positional Plus-Minus in the NBA

Basketball
111280

1. Introduction

How to objectively evaluate player performance is a core question in the field of sports analytics. So far, scholars have proposed a variety of metrics to study player and team performance (Baghal 2012; Franks et al., 2015). Among these, the plus-minus rating is perhaps the most popular one (Engelmann 2011; Fellingham 2022; Hvattum 2019; Macdonald 2012; Rosenbaum 2004; Sill 2010; Winston 2012). In general, plus-minus intends to measure how players contribute to team offense and defense while they are on the court (Grassetti et al., 2021; Sabin 2021). If players are effective in helping teams win, their plus-minus should be high. Due to the nature of sport competition, the plus-minus metric is often divided into two parts, offensive plus-minus and defensive plus-minus. Offensive plus-minus measures players' contributions to team offense, while defensive plus-minus quantifies players' contributions to team defense (Macdonald 2011).

In this study, we explore NBA player plus-minus ratings at the position level by introducing a hierarchical structure to previous plus-minus models (Sabin 2021; Thomas et al., 2013; Yurko et al., 2019). Prior studies of plus-minus predominately focus on estimating plus-minus for individual players (Deshpande and Jensen 2016). For example, Plus-minus metrics such as regularized adjusted plus-minus (Engelmann 2011; Sill 2010), adjusted plus-minus (Rosenbaum 2004), ESPN's real plus-minus (Ilardi and Engelmann 2014), and estimated plus-minus (Snarr 2022), have developed over the past decade to measure player efficiency.

Position level analysis in basketball often faces the challenge of clearly defining player roles on the court (Page et al., 2013; Beckler et al., 2013). Traditionally, five positions are used in basketball, center, power forward, small forward, shooting guard, and point guard (Anıl Duman et al., 2021; McBasketball 2017;). Yet, with the evolvement of basketball games, the definition of player positions becomes less clear, as players with the same position can play vastly different roles on the field (Whitehead 2019). For example, some point guards may focus on orchestrating offense while others may focus on scoring. To better evaluate positional plus-minus, we redefine seven offensive and three defensive positions in this study by using the K-means++ algorithm as well as play type and player position data (Mallepalle et al., 2020).

Overall, we adopt a hierarchical Bayesian linear model to estimate positional offensive and defensive plus-minus in the NBA from 2015-16 to 2021-22 based on redefined player positions. We then compare estimated plus-minus ratings across positions and analyze how these positional values have changed over the years. Understanding positional values provides significant benefits to NBA teams in trades and free agency as these values can serve as replacement levels in player evaluation (Whitehead 2019). For example, NBA teams may compare individual players' plus-minus to the positional plus-minus estimated from this study in order to more accurately assess player efficiency. Studying positional values over multiple seasons also allows teams to observe the overall trend of player efficiency at different positions.

In so doing, our study offers several contributions. First, we present a novel perspective on the values of positions in basketball. Instead of clustering players based on traditional positions, we

rely on play type data to inform players' offensive roles on the field (Whitehead 2019). Then, the plus-minus framework with a hierarchal linear model is employed to derive positional values. This modeling structure takes the performance of teammates and opponents into account while maintaining the flexibility of constructing group level parameters (Sabin 2021). Second, we adopt a Bayesian approach to estimate positional plus-minus. Bayesian analysis allows us to quantify the uncertainty of group parameters so inference can be done easily compared to frequentist methods (Jensen et al., 2009; Santos-Fernandez et al., 2019; Thomas et al., 2013).

The rest of this paper is organized as follows. First, we introduce data and models in sections 2 and 3. We then report the main findings in section 4. Section 5 discusses possible explanations for these findings. We perform additional analyses and report analysis results in section 6. Section 7 concludes this paper with a discussion of limitations and future studies.

2. Data

Our analysis of positional plus-minus in the NBA uses data from three main sources. First, we employ NBA play type data to estimate offensive positions and utilize official player position data to identify defensive positions. Play type data document how players use their offensive possessions (McBasketball 2017). Specifically, the company Synergy Sports classifies each offensive possession into one of the 11 play types. In Table 1, we list the 11 defined play types with detailed descriptions. We calculate the percentage of play types for each player in order to make the numbers comparable across the league. Play type data are available on the NBA official statistics website since the 2015-16 season.

Table 1. Definitions of play types.

Play Type	Description
Cut	Players using cut actions to score
Miscellaneous	Possessions do not fit in any other play types
Post-up	Players receiving the ball in the post to score
P&R Man	Pick and Roll actions with screeners as scorers
Putback	Scoring after receiving offensive rebounds
Transition	Scoring in transition when defense is not set.
Hand-off	Scoring through hand-off actions
Off-screen	Players running off of screens to score
P&R Handler	Pick and Roll actions with ball handlers as scorers
Spot-up	Players spotting up to score
Isolation	One-on-one actions to score

It is important to note that Synergy does not report play types for all players on the NBA website. To be qualified, a player must play at least 10 minutes per game and have at least 10 possessions in a particular play type. For example, if a player played five minutes per game throughout a season, this player would not appear in any play type data set. In other cases, if a player averaged more than 10 minutes per game but played fewer than 10 possessions in one play type, this player would not appear in that particular play type data set. Therefore, there exist some missing values in play type data.

Table 2. Number of players by position and season.

Position	2015-16	2016-17	2017-18	2018-19	2020-21	2021-22	Total
Roll and Cut Big	60	55	59	54	58	64	350
Post-up Big	55	44	39	39	28	20	225
Skilled Big	43	47	52	56	70	76	344
Movement Shooter	35	45	47	49	36	34	246
Spot-up Shooter	69	63	81	88	120	144	565
Versatile Scorer	41	41	45	37	27	34	225
Pick and Roll Attacker	92	94	98	105	105	99	593
Big	98	92	88	87	88	91	544
Forward	126	122	129	137	141	148	803
Guard	171	175	204	204	215	232	1201
Low Usage	81	97	120	103	97	134	632

We take a few steps to process the missing values to ensure the completeness of the data set. First, for players who do not play over 10 minutes per game and thus do not have any play type information, we still include these players in analysis but did not assign any offensive and defensive positions to them. We label players who do not have any play type information as low usage players. Table 2 shows the number of low usage players from 2015-16 to 2021-22. On average, low usage players account for 20 percent of all players in a given season. Second, for players missing particular play type data due to the insufficient number of possessions, we assume equal percentages for these missing play types. For example, if a player misses data on 3 play types with 3 percent of unassigned offensive possessions, then we assume each missing play type accounts for 1 percent of the player's offensive possessions.

Table 3. Defensive position map.

Defensive Position	NBA Official Player Position
Big	C, C-F, F-C
Forward	F, F-G
Guard	G, G-F

After collecting and cleaning play type data, we then draw official player position information from the NBA official statistics website in order to identify players' defensive positions. NBA players may play different positions throughout a season. Thus, it is common to see multiple positions associated with one player. The NBA offers seven combinations of player positions between C (Center), F(Forward), and G(Guard) (see Table 3). We rely on these official player positions to categorize players into one of the three defensive positions, Big, Forward, and Guard.

Big players are often interior defenders whose main responsibilities include protecting the paint area. Guards and Forwards tend to be shorter but more mobile, and thus are good candidates for exterior defenders. We depend on players' primary position, which is often listed in the first letter of official player positions, to determine players' defensive roles. Specifically, If the primary position of a player is C, then this player will be placed into the Big group. Players with the primary position F and G will be assigned to the Forward and Guard group respectively. The only exception is players with the official position F-C. We place this set of players in the Big group even though their primary position is F. This is because F-C players may appear more often as interior defenders with the secondary position being C than regular F players.

Lastly, we gather NBA play-by-play data between 2015-16 to 2021-22 from the NBA official statistics website. We exclude the NBA 2019-20 data as the season was interrupted by the COVID-19 pandemic and had many games played in a neutral site. NBA play-by-play data track a sequence of game related events on the court. They document detailed information such as team scores, time left, and player or team actions at a particular point in an NBA game. Many advanced player metrics are derived from play-by-play data (Engelmann 2017; Franks et al., 2016; Pelechrinis and Papalexakis 2016).

Table 4. Number of stints by season.

Season	Number
2015-16	61502
2016-17	60902
2017-18	60548
2018-19	62133
2020-21	52784
2021-22	61271

From NBA play-by-play data, we extract information related to the performance of stints. A stint can be defined as a unique matchup between five offensive players from one team and five defensive players from another team. We treat a stint as the unit of observation in our statistical model. For each stint in a season, we count the total number of offensive possessions played and the total number of points scored by offensive teams. Table 4 reports the total number of unique stints in each season from 2015-16 to 2021-22. It is important to note that some plus-minus studies do not differentiate offense and defense in stints. Thus, their counts of stints can be much smaller than what we show here (Deshpande and Jensen 2016; Jacobs 2019).

3. Methodology

Our analysis of positional plus-minus in the NBA consists of two steps. First, we determine players' offensive positions by using NBA play type data and the K-means++ clustering algorithm. Then we construct a hierarchical linear model to estimate positional offensive and defensive plus-minus.

We apply K-means++ to cluster players into distinct offensive positions based on players' play types. The K-means++ clustering method is a powerful algorithm for partitioning data points into smaller groups. Unlike the traditional K-means algorithm that selects random data points as initial centroids, K-means++ chooses initial values of centroids that are well separated from each other,

and thus may produce more accurate clustering results (Arthur and Vassilvitskii 2007). Before applying the K-means++ algorithm, we also standardize all the percentages of play types and remove data from the miscellaneous category, which according to the company Synergy Sports, contains all the offensive possessions that cannot be clearly classified into any other play types. In total, we use the standardized percentages of ten different play types in the K-means++ algorithm.

After assigning one offensive and one defensive position to each player, we propose a hierarchical linear model to estimate positional offensive and defensive plus-minus in the NBA. The model has the following specification:

$$Y = X\beta + Z\gamma + e \tag{1}$$

where X is an m by 2n player matrix filled with values of 1, -1, and 0, where n equals the number of unique players in a season. The first n columns represent offensive players while the last n columns indicate defensive players. m denotes the total number of stints in a given season. e is a vector of independent errors with a mean equal to 0 and variance equal to σ^2.

The value of 1 in X indicates a player is on the floor and is with the team playing offense in a stint. -1 shows a player is on the court and is with the team playing defense. 0 means the player is not on the floor. Z is an m by 2k team matrix filled with 1, -1, and 0, where k equals the number of unique teams in a season. The first k columns represent offensive teams and the last k columns denote defensive teams. Similar to the player matrix, the values of 1, -1, and 0 in the Z matrix indicate offensive, defensive, and other teams respectively in a stint.

The response variable Y is an m by 1 matrix, representing points scored per 100 possessions by the offensive team in a stint. For example, if the offensive team in a stint scored 20 points in 20 possessions, y will be equal to 100. Let y_i denote the ith row of Y. Model (2) implies that y_is are independent of each other and follow a normal distribution with constant variance, conditional on X^i and Z^i,

$$y_i|X^i, Z^i \sim N(X^i\beta + Z^i\gamma, \sigma^2) \tag{2}$$

β is a vector of 2n player effects. If a player is assigned to a position group j, we assume the player effect β_{p_j} follows a normal distribution with a group mean of μ_j and group variance of τ_j^2, where j is from 1 to 10 with 7 offensive and 3 defensive positions derived in cluster analysis. p_j is a subset of {1, ..., 2n} that belongs to position j. If a player is not associated with any group, (i.e., a low usage player), then we assume its player effect β_p follows a normal distribution with a mean equal to 0 and variance equal to 100, where p refers to a low usage player. In summary, we put the following priors on model coefficients β.

$$\begin{aligned} \beta_{p_j} &\sim N(\mu_j, \tau_j^2) \\ \beta_p &\sim N(0,100) \end{aligned} \tag{3}$$

γ is a vector of team effects. We assume the team effect γ_t comes from a normal distribution with a mean equal to 0 and variance equal to 100.

$$\gamma_t \sim N(0, 100) \tag{4}$$

where t represents an individual team.

Below are the hyperprior distributions for the group mean μ_j, group variance τ_j^2, error variance σ^2.

$$\mu_j \sim N(0, 100)$$
$$\tau_j^2 \sim IG(0.01, 0.01) \tag{5}$$
$$\sigma^2 \sim IG(0.01, 0.01)$$

We then construct a Gibbs sampler to derive the posterior distributions for group level parameters (i.e., μ_j, and τ_j^2). The position level parameters are key of interest, but similar inference can be done for individual level and team level parameters in this study. The details of the Gibbs sampling can be found in Appendix A.

4. Results

4.1 Player positions

We apply the K-means++ algorithm to play type data to derive seven clusters in this study with each cluster representing one offensive player position. The elbow plot (see Figure 1) shows the number of clusters and the corresponding total within-clusters sum of squares. Choosing seven clusters is based on the analysis of the K-means++ elbow plot and domain knowledge in basketball. Since the K-means clustering method is a heuristic algorithm, it will be difficult to determine the best number of clusters by just examining K-means clustering results (Brusco and Cradit 2001). Therefore, scholars often search for the stopping point where the subsequent number of clusters no longer significantly reduces the total within-clusters sum of squares. We also consider if the chosen number of clusters will result in a sufficient number of players in each cluster. Overall, we select seven offensive positions for this analysis as the optimal number of clusters.
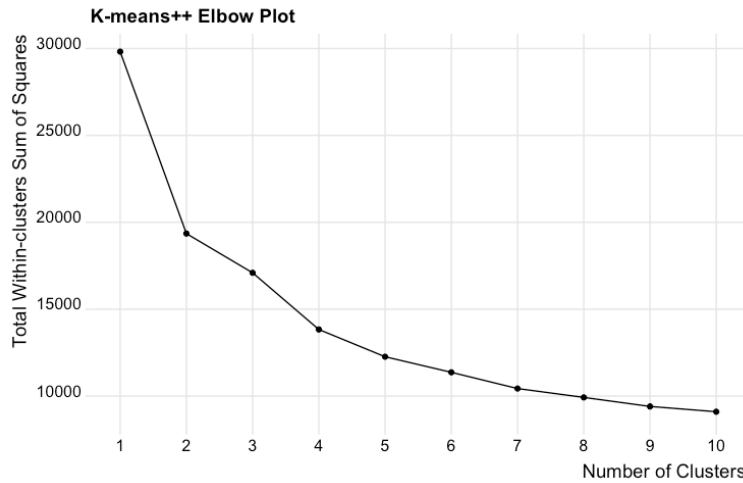

Figure 1. K-means++ elbow plot

The centroid of clustered offensive positions can be found in Table 5. Based on the centroid values of clusters, we rely on basketball domain knowledge to name derived positions, Roll and Cut Big, Post-up Big, Skilled Big, Movement Shooter, Spot-up Shooter, Versatile Scorer, and Pick and Roll Attacker. Specifically, Roll and Cut Bigs tend to use pick and roll as well as cut actions to score points. They also contribute to team offense through putbacks. Post-up Bigs heavily rely on post-up actions to score. They also utilize pick and roll actions to score but not as frequently as Roll and Cut

Bigs. It is interesting to note that the number of Post-up Bigs in the league has significantly reduced over time, from 55 players in 2015-16 to only 20 players in 2021-22 (see Table 2). Skilled Bigs are players who not only score through pick and rolls, but also are capable of spot-up shooting.

Table 5. Centroid percentages of play types for each offensive position.

	Cut	Post-up	P&R Man	Putback	Transition	Hand-off	Off-screen	P&R Handler	Spot-up	Isolation
Roll and Cut Big	24%	6%	22%	18%	8%	1%	1%	1%	8%	2%
Post-up Big	12%	23%	19%	10%	8%	1%	2%	1%	13%	5%
Skilled Big	9%	6%	12%	7%	12%	3%	3%	3%	36%	4%
Movement Shooter	4%	2%	1%	2%	16%	11%	15%	14%	27%	3%
Spot-up Shooter	7%	2%	3%	4%	21%	5%	4%	10%	35%	4%
Versatile Scorer	3%	6%	2%	3%	15%	4%	4%	26%	15%	16%
Pick and Roll Attacker	3%	2%	1%	2%	15%	6%	3%	36%	20%	7%

According to Table 5, Movement Shooters seem to be the best shooters in the league as their offensive possessions largely come from off-screen, hand-off, and spot-up actions. Scoring through these actions may require strong shooting skills. Unlike Movement Shooters, Spot-up Shooters exceedingly rely on spot-up shooting. Often, spot-up shooters will locate in the corner of a basketball court and wait to catch and shoot the ball. Interestingly, the number of Spot-up Shooters has increased considerably since the 2020-21 season (see Table 2). Versatile Scorers are all-around players who are capable of scoring through multiple actions such as isolation, transition, and pick and roll. Compared to Versatile Scorers, Pick and Roll Attackers predominately employ pick and roll actions to score points and spend a significantly lower number of possessions on isolation actions.

After we determine offensive and defensive positions for players, we run the same hierarchical Bayesian linear model for each of the six NBA seasons separately to estimate positional plus-minus. Tables 6 and 7 summarize the mean and standard deviation of position level mean and variance parameters. In the following text, we focus on comparing the values of parameters across positions and seasons as well as interpreting the differences of the values.

Table 6. Mean and standard deviation of positional plus-minus mean estimate.

Position	2015-16	2016-17	2017-18	2018-19	2020-21	2021-22
Roll and Cut Big	10 (1.4)	10.8 (1.4)	9.4 (1.2)	13.2 (1.4)	12.3 (1.2)	14.3 (1.4)
Post-up Big	12.4 (1.3)	12.5 (1.4)	11.3 (1.5)	14.6 (1.4)	11.8 (1.5)	13.6 (1.6)
Movement Shooter	15.3 (1.4)	14.1 (1.3)	14.9 (1.1)	13.2 (1.2)	15.6 (1.2)	14.2 (1.5)
Spot-up Shooter	13.6 (1.3)	13.1 (1.1)	12.7 (1)	13.2 (1.2)	12.9 (0.9)	13.3 (1.3)
Versatile Scorer	16.8 (1.3)	15.7 (1.3)	14.7 (1.2)	14.9 (1.4)	17.8 (1.2)	17.6 (1.5)
Skilled Big	11.9 (1.4)	12.4 (1.2)	12.5 (1.2)	12 (1.3)	12.7 (1)	13.6 (1.3)
Pick and Roll Attacker	15.2 (1.3)	13.7 (1.2)	13.7 (1)	13 (1.2)	13.4 (1)	12.6 (1.3)

Big	-4.6 (1.3)	-4.9 (1.1)	-6.1 (1.1)	-4.9 (1.1)	-6.1 (1.2)	-5.8 (1.4)
Forward	-6.3 (1.2)	-6.9 (1.1)	-7.7 (0.9)	-7.6 (1.1)	-7.4 (1)	-7.5 (1.3)
Guard	-8.1 (1.2)	-9.3 (1.1)	-8.6 (1)	-9.3 (1.1)	-9.5 (1)	-7.9 (1.3)

Note. Standard deviations are in parentheses.

Table 7. Mean and standard deviation of positional plus-minus variance estimate.

Position	2015-16	2016-17	2017-18	2018-19	2020-21	2021-22
Roll and Cut Big	3.3 (4)	0.7 (1.2)	1.6 (2.7)	4.7 (4.8)	3.3 (3.9)	4.7 (4.6)
Post-up Big	2.4 (3.1)	19.9 (8.5)	28.9 (12.4)	11.7 (7.1)	14.4 (11.6)	11 (10.1)
Movement Shooter	8.2 (6.2)	12.7 (7.1)	1.8 (2.5)	1.9 (2.8)	6.3 (5.8)	4.3 (4.9)
Spot-up Shooter	1.1 (1.6)	0.4 (0.8)	0.5 (0.9)	9.5 (4.6)	1.1 (1.5)	1.1 (1.5)
Versatile Scorer	6.7 (5.5)	12.9 (7.5)	15 (7.5)	13.3 (8)	1.7 (3)	12.6 (7.6)
Skilled Big	7.2 (5.9)	1 (1.9)	6.8 (6.2)	2 (3)	1.2 (2.3)	0.8 (1.3)
Pick and Roll Attacker	1.8 (2.2)	10.9 (3.9)	1.6 (1.9)	7 (3.5)	3.5 (2.9)	11.4 (4.3)
Big	4.3 (3.3)	2.5 (2.7)	1.9 (2.5)	1.2 (1.8)	1.2 (2)	1.3 (1.8)
Forward	3.4 (2.7)	4.1 (2.6)	1.6 (1.9)	4.8 (2.5)	1.7 (2.2)	1.8 (1.9)
Guard	5.7 (2.2)	2.7 (1.9)	7.6 (2.5)	2.3 (1.9)	4.7 (2.4)	5.3 (2)

Note. Standard deviations are in parentheses.

4.2 Comparison of plus-minus across positions by season

Figure 2 visualizes estimated positional offensive plus-minus from the NBA 2015-16 to 2021-22 seasons.[1] The figure shows that Versatile Scorers are consistently ranked higher than the other positions in offensive plus-minus. To estimate the probabilities of Versatile Scorers' offensive plus-minus higher than the other positions' offensive plus-minus, we use samples from the posterior distributions of offensive plus-minus and calculate the differences between these posterior samples. Figure 3 shows the 80% credible intervals of the differences of offensive plus-minus between Versatile Scorers and the other offensive positions from 2015-16 to 2021-22. It is evident that most of the 80% credible intervals of the differences are above zero.[2] This suggests that Versatile Scorers are the most valuable offensive contributors in most seasons.

---

[1] Dashed lines represent posterior means.
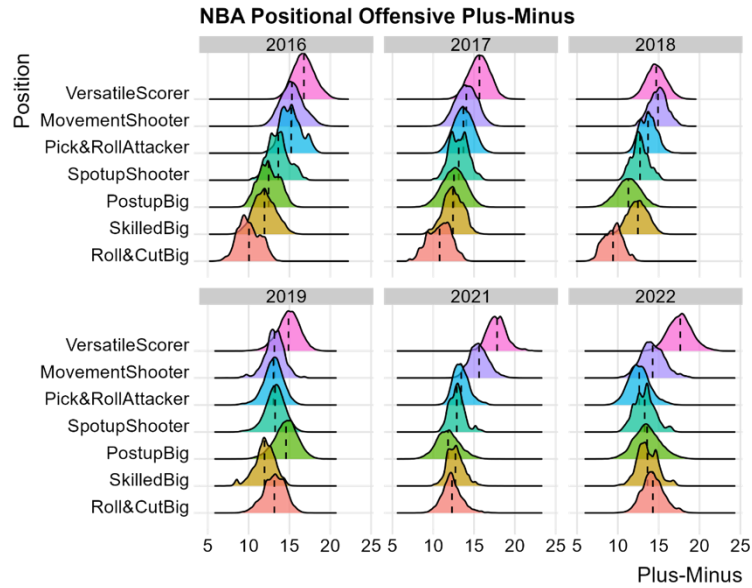[2] Red dots represent the means of differences.

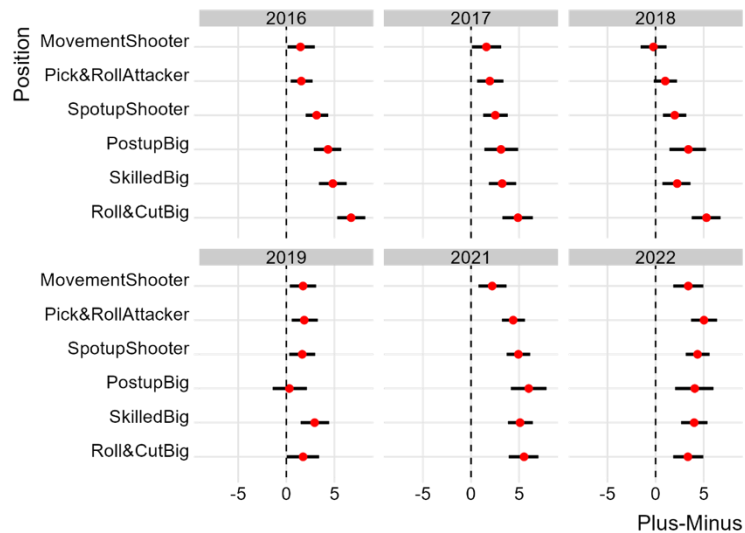Figure 2. NBA positional offensive plus-minus from 2015-16 to 2021-22.

Figure 3. 80% credible intervals of the differences of offensive plus-minus between Versatile Scorers and the other offensive positions.

Another observation from Figure 2 is that the gaps of positional offensive plus-minus have shortened over the years. In particular, the differences of offensive plus-minus in 2015-16 seemed fairly high from position to position, with Versatile Scorers' offensive plus-minus leading and Roll and Cut Bigs' offensive plus-minus at the bottom of the list. The gaps disappeared in 2021-22 when the posterior means of positional offensive plus-minus move closer to each other. The only exception is Versatile Scorers whose offensive plus-minus are constantly higher than the other positions' offensive plus-minus. This implies while the differences of positional offensive plus-minus have become harder to distinguish over time, Versatile Scorers are still the most critical parts of team offense in the NBA.
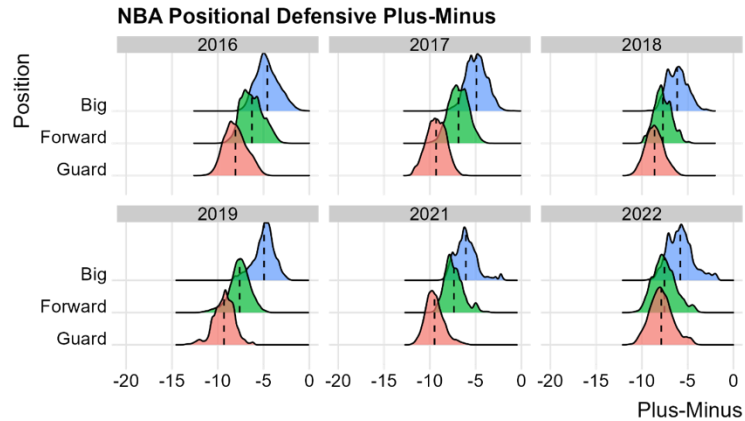
Figure 4. NBA positional defensive plus-minus from 2015-16 to 2021-22.

Moving to positional defensive plus-minus, Figure 4 shows the posterior means of Bigs' defensive plus-minus are consistently h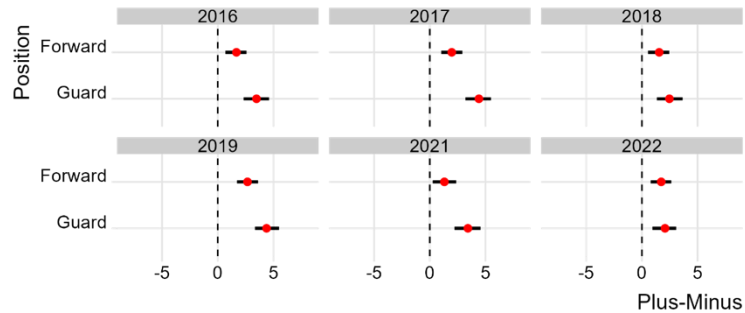igher than those of the other positions' defensive plus-minus. Figure 5 shows that the 80% credible intervals of the differences of defensive plus-minus between Bigs and the other defensive positions are well above zero from 2015-16 to 2021-22, indicating strong evidence of Bigs providing more defensive contributions than the other defensive positions.

Figure 5. 80% credible intervals of the differences of defensive plus-minus between Bigs and the other defensive positions.

Figure 6 shows that the variance of Spot-up Shooters' offensive plus-minus is the lowest among all offensive positions in most seasons. In contrast, the variances of Post-up Bigs' and Versatile Scorers' offensive plus-minus seem to be considerably higher than those of the other positions' offensive plus-minus. This finding suggests that player offensive plus-minus may differ significantly in these two positions, while Spot-up Shooters' offensive efficiency does not seem to vary much from player to player.
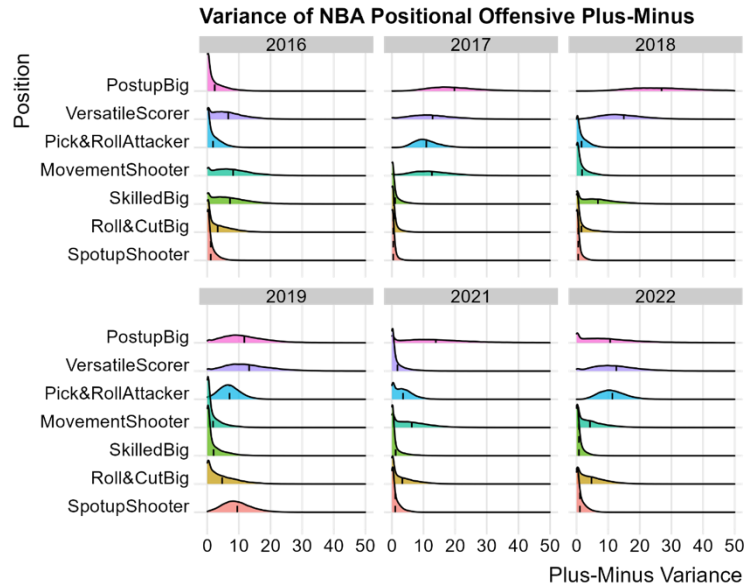
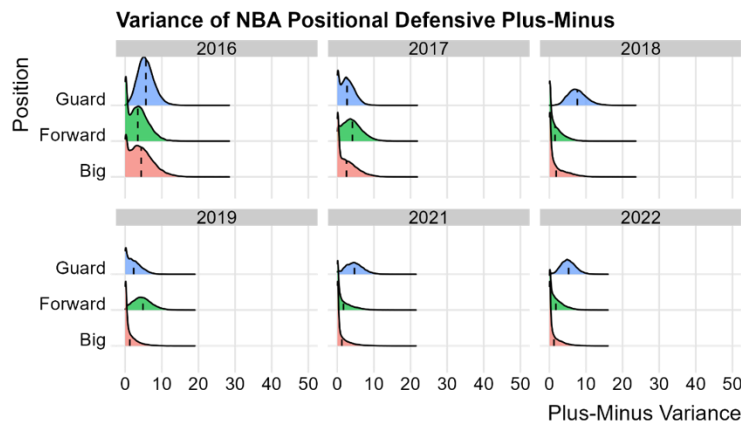Figure 6. Variance of NBA positional offensive plus-minus from 2015-16 to 2021-22.

Figure 7. Variance of NBA positional defensive plus-minus from 2015-16 to 2021-22.

Figure 7 displays the posterior distributions of the variance of positional defensive plus-minus. It appears that the variance of Guards' defensive plus-minus is much higher than those of the other two positions' defensive plus-minus. This indicates that Guard players tend to have higher variability of defensive efficiency, relative to Forwards and Bigs. The 80% credible intervals of the variance differences between Guards and the other defensive positions do not include zero in most cases.[3] This suggests there is some evidence that the variance of Guards' defensive plus-minus is relatively higher.

4.3 Comparison of plus-minus across seasons by position
Figure 8 visualizes the posterior distributions of offensive plus-minus across seasons by position. There seems to be a notable improvement for Roll and Cut Bigs' offensive plus-minus as their posterior means increased significantly from 2015-16 to 2021-22. Post-up Bigs and Skilled Bigs

---

[3] 80% credible intervals are not shown here. They are available upon request.

also experienced some progress on offensive plus-minus over the years. Other positions' offensive plus-minus stays relatively stable from season to season.
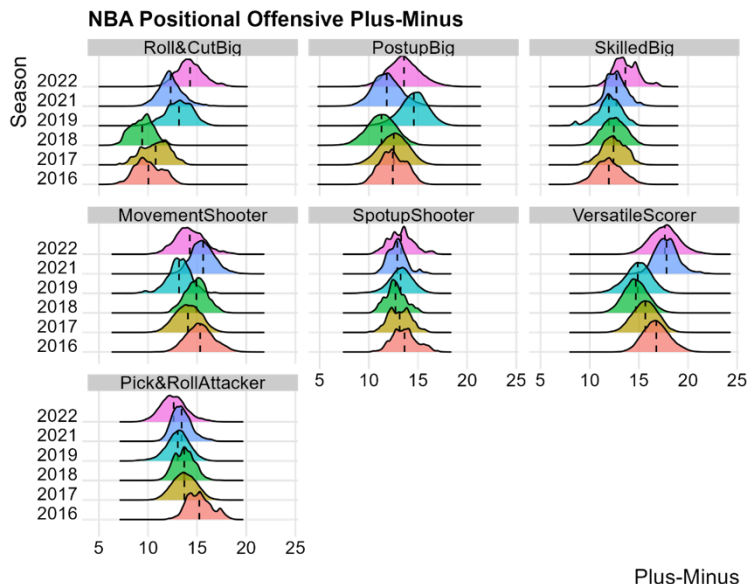

Figure 8. NBA offensive plus-minus by position.

Figure 9 shows the posterior distributions of the variance of offensive plus-minus across seven offensive positions. It seems the variances of positional offensive plus-minus are relatively consistent over the seasons. Lastly, we note the posterior distributions of positional defensive plus-minus and the variance of positional defensive plus-minus do not exhibit any pattern over the years.
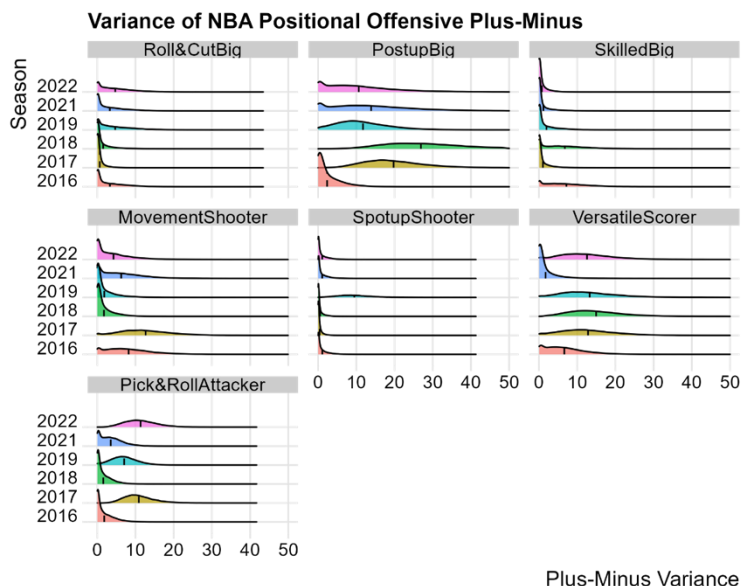

Figure 9. Variance of NBA offensive plus-minus by position.

5. Discussion

Recall in Figures 2 and 3 we show that Versatile Scorers provide the highest offensive value among all offensive positions. We define Versatile Scorers as players who rely on isolation, transition, and pick and roll actions to score points. As such, these players are likely to be the ball-dominant players who control the ball most of the time on the court. Teams often allow star players to be the ball-dominant players as they tend to be more efficient in scoring. Therefore, it is not surprising that Versatile Scorers have higher offensive plus-minus than the other positions.

By comparing results across the six seasons, we show that the gaps of offensive plus-minus between the seven offensive positions have shrunk over time. After further examining how offensive plus-minus have changed over the seasons for each position, we notice that the improvement of Roll and Cut Bigs' offensive efficiency may contribute to the shortened gaps, as the posterior means of their offensive plus-minus increased steadily from 2015-16 to 2021-22. Given the number of Roll and Cut Bigs stays relatively stable over the years (see Table 2), we believe the increased offensive plus-minus may arise from Roll and Cut Bigs being more efficient on offense. We also observe the growth of Post-up Bigs' and Skilled Bigs' offensive plus-minus over time, although their plus-minus ratings do not seem to increase as dramatically as Roll and Cut Bigs' offensive plus-minus.

Our analysis also shows the variance of Spot-up Shooters' offensive plus-minus ranks at the bottom in most seasons from 2015-16 to 2020-21. This may indicate the variation of Spot-up Shooter's offensive plus-minus is relatively small, compared to the other positions. This finding may come from the fact that Spot-up Shooters do not play significant roles in team offense, as their primary responsibility is to make open shots when available. On the contrary, the variances of Post-up Bigs' and Versatile Scorers' offensive plus-minus are much higher than those of the other positions' offensive plus-minus. The steep decline in quantity and the relatively low number of Post-up Bigs in the NBA may partially explain the high variation of Post-up Bigs' offensive plus-minus. In addition, the high variation of offensive plus-minus for Versatile Scorers and Post-up Bigs may suggest that the scoring ability for offensive-oriented positions can vary significantly from one player to another in the NBA.

On the defensive side, we show Bigs tend to provide more defensive values than the other positions. This finding is not surprising as Big players not only need to guard opposing Bigs, but also help stop opponents from scoring in the paint area. Therefore, interior players often carry more defensive responsibility and may have higher defensive impacts than exterior players. Guards, limited by their size, offer the least defensive value to teams. It is also interesting to note that while Guards provide the lowest defensive plus-minus, they have the highest variance of defensive plus-minus among the three defensive positions. In other words, Guards' defensive plus-minus is more spread out from the average value, relative to Bigs' and Forwards' defensive plus-minus. This finding suggests that there may exist some exceptional defenders or some extremely poor defenders in the Guard position.

6. Additional Analysis

To check the robustness of our estimation results, we re-estimate positional plus-minus using stint data extracted from play-by-by without garbage time. We use the definition of garbage time from the basketball analytics website, Cleaning the Glass (Falk, n.d.). In general, the website considers the time when the score differential is large toward the end of the game and two teams no longer field their best players as garbage time. Specifically, three factors are used to determine garbage time,

the quarter, score differential, and the number of starters on the floor. For example, if a game is in the fourth quarter, one team is leading the other team by 15 points with 2 minutes left on the clock, and the total number of starters on the floor combined from the two teams is 1, then the website will consider this period as garbage time. The details of garbage time thresholds can be found in Appendix B.

Overall, we find similar estimation results as before. To better visualize these findings, we place the estimates from play-by-play data with and without garbage time side by side in Figures 10 to 11. It is clear that removing garbage time from play-by-play data does not significantly alter estimated plus-minus, except for the 2020-21 season when positional plus-minus from the full data set seems higher than positional plus-minus derived from the data set without garbage time. We suspect this is because 2020-21 season games contain more garbage time than games in the other seasons, as 2020-21 is the first full season since the interrupted 2019-20 season due to the COVID-19 pandemic. Data analysis provides some support for this possibility. After counting the percentage of garbage time in each of the six seasons examined in this study, it appears that almost 4 percent of possessions are classified as garbage time in the 2020-21 season, whereas only 3.4 to 3.7 percent of possessions are considered garbage time in the other seasons.
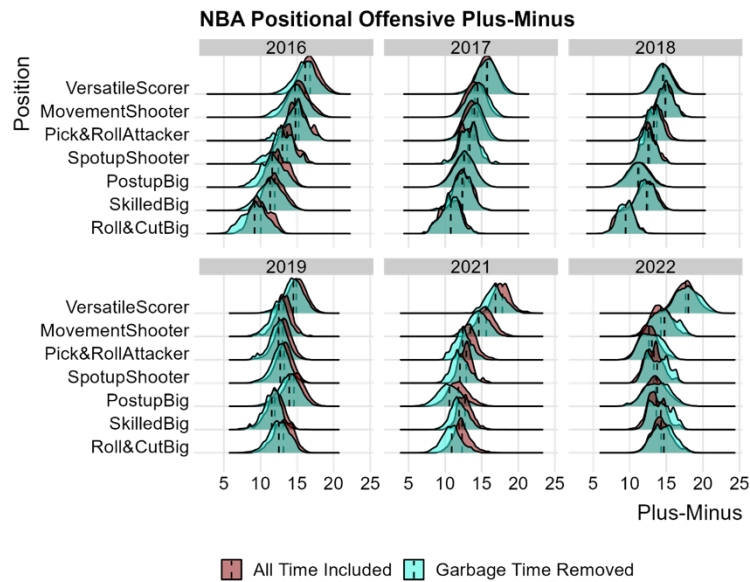

Figure 10. Comparison of offensive plus-minus across seasons with and without garbage time.
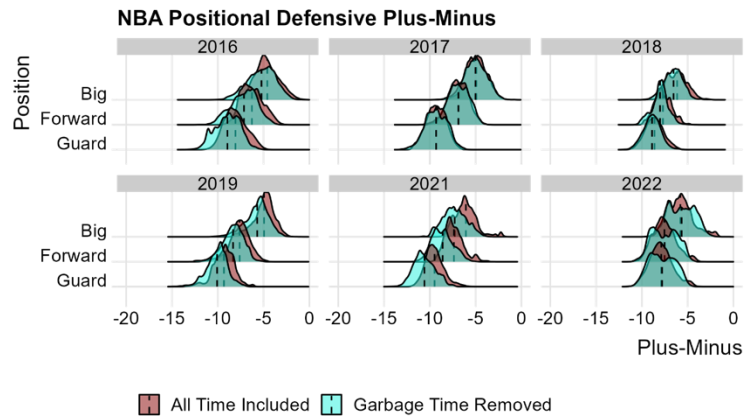

Figure 11. Comparison of defensive plus-minus across seasons with and without garbage time.

7. Conclusion

In this study, we have redefined seven offensive positions and three defensive positions in the NBA and employed a hierarchical Bayesian linear model to estimate positional plus-minus from 2015-16 to 2021-22 based on these offensive and defensive positions. Our analysis shows that Versatile Scorers and Bigs have the highest positional offensive and defensive plus-minus respectively throughout the seasons. These are not surprising results, given the fact that Versatile Scorers tend to be ball-dominant and star players on NBA teams. Big players provide valuable defensive contributions as they not only guard opposing Bigs but also help protect the paint area. We also show that the variance of Spot-up Shooters' offensive plus-minus is the lowest among the seven offensive positions, while the variances of Post-up Bigs' and Versatile Scorers' offensive plus-minus are substantially higher. On the defensive side, the variance of Guards' defensive plus-minus is the highest among the three defensive positions.

By examining positional offensive plus-minus from 2015-16 to 2021-22, we have noted that the gaps of offensive plus-minus between the seven offensive positions have shortened over time. That is, the values of positional offensive plus-minus have become closer over time, except for Versatile Scorers whose offensive plus-minus is consistently ranked higher than the other positions' offensive plus-minus. We have also highlighted that the rise of Roll and Cut Bigs' offensive plus-minus may significantly contribute to the shortened gaps.

Our study of positional offensive and defensive plus-minus has a few limitations. First, the response variable, points scored per 100 possessions, employed in this study does not reflect the importance of points in a game. In fact, we treat points scored or allowed equally throughout a game. However, it is common to see points scored in a close game toward the end of the game are far more important than points acquired in the first quarter. Second, the clustering results for offensive positions are subjective as the K-means++ algorithm is a heuristic method. Therefore, the seven offensive positions we proposed in this study may not be the best solution for grouping players.

Future studies of positional plus-minus may consider using win probability as the response variable. The calculation of win probability takes game situations into consideration. Thus, regressing the change in win probability over stints may better estimate positional contributions to team success (Deshpande and Jensen 2016). Future research may also employ new data to cluster players. In particular, the availability of player tracking data in major sports leagues may allow scholars to produce more detailed and accurate classifications of player positions in sports.

Appendix A

We construct a hierarchical Bayesian liner model to estimate positional plus-minus in this study. The Gibbs sampling method with 30,000 iterations and a burn-in period of 2,000 is used to derive the estimates in the model. Our variables of interest are the mean and variance of positional plus-minus μ_j and τ_j^2, where j is from 1 to 10 with 7 offensive positions and 3 defensive positions.

Since in Equation (5) we assume μ_j follows a normal distribution with a mean equal to 0 and variance equal to 100 and τ_j^2 follows an inverse gamma distribution with both the shape and scale equal to 0.01, we can derive the following conditional posterior distributions for μ_j and τ_j^2.

$$p(\mu_j|.) \propto N(\left(k + \frac{\tau_j^2}{100}\right)^{-1} \left(\sum \beta_{p_j}\right), \tau_j^2 \left(k + \frac{\tau_j^2}{100}\right)^{-1}) \tag{6}$$

$$p(\tau_j^2|.) \propto IG\ (\frac{k}{2} + 0.01, 0.5 * \sum (\beta_{p_j} - \mu_j)^2 + 0.01) \tag{7}$$

where k is the number of players that belong to group j. p_j is a subset of {1, ..., 2n} that belongs to group j. β represents player effects.

Appendix B

Table 8. Thresholds of NBA garbage time defined by the website Cleaning the Glass.

	Condition
1	Game is in the 4th quarter
2	Score differential is larger than 25 when 9 to 12 minutes left
3	Score differential is larger than 20 when 6 to 9 minutes left
4	Score differential is larger than 10 when 0 to 6 minutes left
5	A total of two or fewer starters combined from the two teams on the field

The website also notes the game can never change from being garbage time to being non-garbage time. If any of the above conditions is not met and the game switches from being garbage time to being non-garbage time, then the garbage time clock resets.

References

Anıl Duman, Eyüp, Bahar Sennaroğlu, and Gülfem Tuzkaya. 2021. "A cluster analysis of basketball players for each of the five traditionally defined positions." *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*. https://doi.org/10.1177/17543371211062064.

Arthur, David, and Sergei Vassilvitskii. 2006. "*k-means++: The advantages of careful seeding*." Stanford.

Baghal, Tarek. 2012. "Are the" four factors" indicators of one factor? an application of structural equation modeling methodology to NBA data in prediction of winning percentage." *Journal of Quantitative Analysis in Sports* 8, no. 1. https://doi.org/10.1515/1559-0410.1355.

Beckler, Matthew, Hongfei Wang, and Michael Papamichael. 2013. "NBA oracle." *Zuletzt besucht am* 17, no. 20082009.9.

Brusco, Michael J., and J. Dennis Cradit. 2001. "A variable-selection heuristic for K-means clustering." *Psychometrika* 66, no. 2: 249-270. https://doi.org/10.1007/BF02294838.

Deshpande, Sameer K., and Shane T. Jensen. 2016. "Estimating an NBA player's impact on his team's chances of winning." *Journal of Quantitative Analysis in Sports* 12, no. 2: 51-72. https://doi.org/10.1515/jqas-2015-0027.

Engelmann, Jeremias. 2011. "A new player evaluation technique for players of the National Basketball Association (NBA)." In *Proceedings of the MIT Sloan Sports Analytics Conference*.

Engelmann, Jeremias. 2017. "Possession-based player performance analysis in basketball (adjusted+/–and related concepts)." In *Handbook of statistical methods and analyses in sports*, 231-244. Chapman and Hall/CRC.

Falk, Ben. n.d. "Garbage Time." https://cleaningtheglass.com/stats/guide/garbage_time.

Fellingham, Gilbert W. 2022. "Evaluating the performance of elite level volleyball players." *Journal of Quantitative Analysis in Sports* 18, no. 1: 15-34. https://doi.org/10.1515/jqas-2021-0056.

Franks, Alexander M., Alexander D'Amour, Daniel Cervone, and Luke Bornn. 2016. "Meta-analytics: tools for understanding the statistical properties of sports metrics." *Journal of Quantitative Analysis in Sports* 12, no. 4: 151-165. https://doi.org/10.1515/jqas-2016-0098.

Franks, Alexander, Andrew Miller, Luke Bornn, and Kirk Goldsberry. 2015. "Counterpoints: Advanced defensive metrics for nba basketball." In *9th annual MIT sloan sports analytics conference, Boston, MA*.

Grassetti, Luca, Ruggero Bellio, Luca Di Gaspero, Giovanni Fonseca, and Paolo Vidoni. 2021. "An extended regularized adjusted plus-minus analysis for lineup management in basketball using play-by-play data." *IMA Journal of Management Mathematics* 32, no. 4: 385-409. https://doi.org/10.1093/imaman/dpaa022.

Hvattum, Lars Magnus. 2019. "A comprehensive review of plus-minus ratings for evaluating individual players in team sports." *International Journal of Computer Science in Sport* 18, no. 1: 1-23. https://doi.org/10.2478/ijcss-2019-0001.

Ilardi, Steve., and Engelmann, Jeremias. 2014. "The Next Big-Thing: Real Plus-Minus." https://www.espn.com/nba/story/_/id/10740818/introducing-real-plus-minus.

Jacobs, Justin. 2019. "Exercising Error: Quantifying Statistical Tests Under RAPM (Part IV)." https://squared2020.com/2019/10/03/exercising-error-quantifying-statistical-tests-under-rapm-part-iv/.

Jensen, Shane T., Blakeley B. McShane, and Abraham J. Wyner. 2009. "Hierarchical Bayesian modeling of hitting performance in baseball." *Bayesian Analysis* 4, no. 4: 631-652. https://doi.org/10.1214/09-BA424.

Macdonald, Brian. 2011. "A Regression-Based Adjusted Plus-Minus Statistic for NHL players." *Journal of Quantitative Analysis in Sports* 7 no. 3. https://doi.org/10.2202/1559-0410.1284.

Macdonald, Brian. 2012. "Adjusted plus-minus for NHL players using ridge regression with goals, shots, Fenwick, and Corsi." *Journal of Quantitative Analysis in Sports* 8, no. 3. https://doi.org/10.1515/1559-0410.1447.

Mallepalle, Sarah, Ronald Yurko, Konstantinos Pelechrinis, and Samuel L. Ventura. 2020. "Extracting NFL Tracking Data from Images to Evaluate Quarterbacks and pass Defenses." *Journal of Quantitative Analysis in Sports* 16, no2: 95–120. https://doi.org/10.1515/jqas-2019-0052.

McBasketball, Cranjis. 2017. "Nylon Calculus: How to understand Synergy play type categories." https://fansided.com/2017/09/08/nylon-calculus-understanding-synergy-play-type-data/.

Page, Garritt L., Bradley J. Barney, and Aaron T. McGuire. 2013. "Effect of position, usage rate, and per game minutes played on nba player production curves." *Journal of Quantitative Analysis in Sports* 9, no. 4: 337-345. https://doi.org/10.1515/jqas-2012-0023.

Pelechrinis, Konstantinos, and Evangelos Papalexakis. 2016. "The anatomy of american football: evidence from 7 years of NFL game data." *PloS one* 11, no. 12: e0168716. https://doi.org/10.1371/journal.pone.0168716.

Rosenbaum, T. Dan. 2004. "Measuring How NBA Players Help Their Teams Win." http://www.82games.com/comm30.htm#_ftn1.

Sabin, R. Paul. 2021. "Estimating player value in American football using plus–minus models." *Journal of Quantitative Analysis in Sports* 17, no. 4: 313-364. https://doi.org/10.1515/jqas-2020-0033.

Santos-Fernandez, Edgar, Paul Wu, and Kerrie L. Mengersen. 2019. "Bayesian statistics meets sports: a comprehensive review." *Journal of Quantitative Analysis in Sports* 15, no. 4: 289-312. https://doi.org/10.1515/jqas-2018-0106

Sill, Joseph. 2010. "Improved NBA adjusted+/-using regularization and out-of-sample testing." In *Proceedings of the 2010 MIT Sloan Sports Analytics Conference.*

Snarr, Taylor. (2022). "NBA player metric comparison." https://dunksandthrees.com/blog/metric-comparison.

Thomas, A. C., Samuel L. Ventura, Shane T. Jensen, and Stephen Ma. 2013. "Competing process hazard function models for player ratings in ice hockey." *The Annals of Applied Statistics* 7, no. 3: 1497-1524.

Whitehead, Todd. (2019). "Nylon Calculus: Grouping players by offensive role, again." https://fansided.com/2019/05/29/nylon-calculus-grouping-players-offensive-role-again/

Winston, Wayne L. 2012. "Mathletics." In *Mathletics.* Princeton University Press.

Yurko, Ronald, Samuel Ventura, and Maksim Horowitz. 2019. "nflWAR: a reproducible method for offensive player evaluation in football." *Journal of Quantitative Analysis in Sports* 15, no. 3: 163-183. https://doi.org/10.1515/jqas-2018-0010.