

# Estimating NBA Team Shot Selection Efficiency from Aggregations of True, Continuous Shot Charts: A Generalized Additive Model Approach

Basketball  
193942

GAM dashboard: <https://sportdataviz.syr.edu/TrueShotChart/>

Open-Source Code Repository: [https://github.com/Syracuse-University-Sport-Analytics/continuous\\_shot\\_selection](https://github.com/Syracuse-University-Sport-Analytics/continuous_shot_selection)

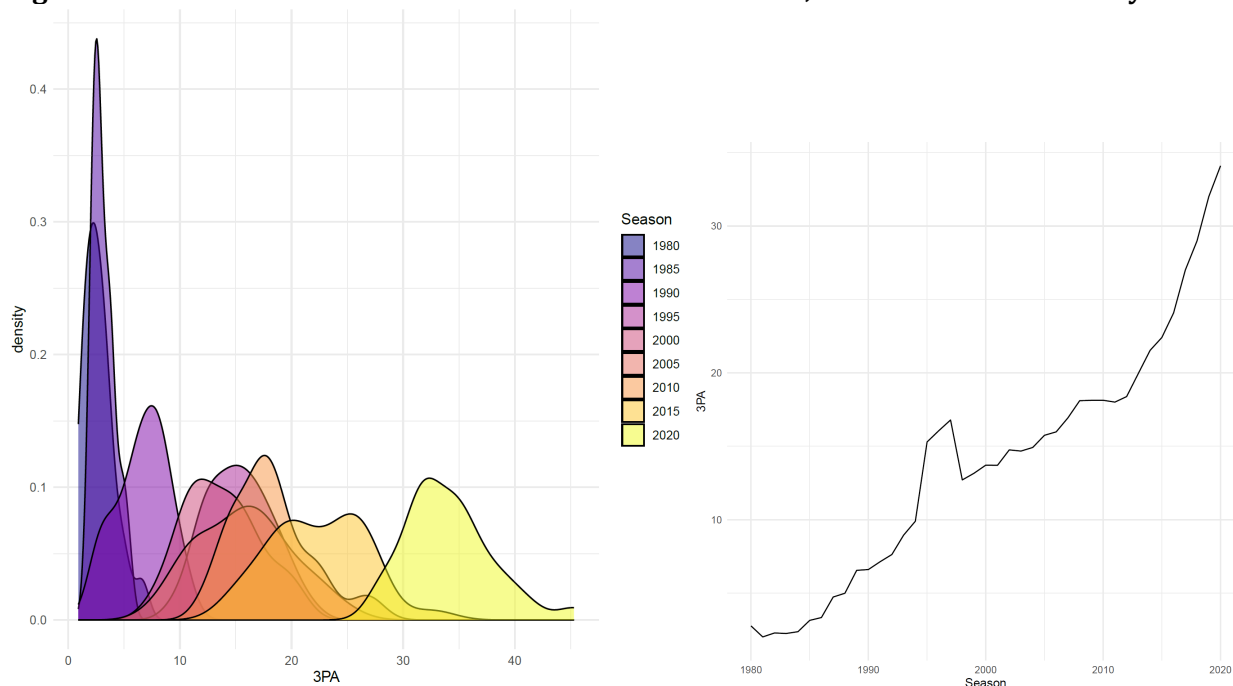
## 1. Introduction

We develop a novel type of basketball shot chart that uses a generalized additive model to estimate total shot proficiency continuously in the half-court. This shot chart incorporates missed shots that draw a shooting foul, and shot-pursuant free throw scoring, to determine total or true scoring yield following a shot decision. A traditional discrete, or binned, shot chart is a size- and color-coded spatial plot that tracks both location-dependent volume and points yielded from the floor, respectively, on field goal attempts (FGAs). Shot charts provide distilled summaries of shot-distributional efficiency and are therefore a leading analytic tool for team game-planning (see, e.g., Papalexakis and Pelechrinis 2018; Jiao, Hu, and Yan 2021; Jieying, Guanyu, and Jun 2021; Franks, Miller, Bornn, and Goldsberry 2015; Fichman and O'Brien 2019; Skinner and Goldman 2015; Goldman and Rao 2011; Narayan 2019; Winston, Nestler, and Pelechrinis 2022). While descriptive, traditional shot charts do not account for free throw scoring pursuant to shots from the field, nor the locations of those shots from the field that lead to the free throw scoring opportunities. By considering missed shots that result in a shooting foul, which are not considered FGAs, and made shots that result in an “and-one” free throw opportunity, we correct for location-conditional scoring yield estimate distortions in traditional shot charts.

Despite their limitations, traditional shot charts have been integral to improving shooting efficiency in the NBA. In a 2020 NPR interview, Kirk Goldsberry, a basketball shot chart pioneer, discussed the ability to infer disequilibria from shot chart aggregations: “It’s that wild margin of inefficiency that’s driven sort of these cartoonish trends and the rapid increase in 3-point shooting across the NBA.” Shot charts, and their aggregations, have served as a key input in the NBA’s three-point revolution. Figures 1a-b show the increased reliance on three-point shooting in the NBA during the past decade, which has allowed teams to further leverage efficiency gains from the three-point premium. The righthand side plot of Figure 1 shows the time trend in average number of three-point attempts (3PAs) per NBA team-game. We observe a marked increase in the locally upward slope of the trend beginning in 2013-14, coinciding with the ubiquitous adoption of SportVU player tracking technology in all NBA arenas. The only precedent for such a steep rise in 3PA volume occurred from 1994 through 1997, when the three-point line was moved inward “above the break” from 23.75 feet to 22 feet before being restored to its previous location. These negating leaguewide three-point line position changes account for the locally steep rise-and-fall pattern of the mid-1990s. The lefthand

side plot of Figure 1 features team-game level three-point attempt distributions for the 1980, 1985, 1990,..., and 2020 seasons (i.e., at five-year staggers). This plot also demonstrates marked rightward shifts in the team-game level 3PA density plots of 2015 and 2020 relative to earlier such plots.

**Figure 1a and 1b: NBA 3-Point Shots Taken Per Team-Game, Time Trend and Density Plots**



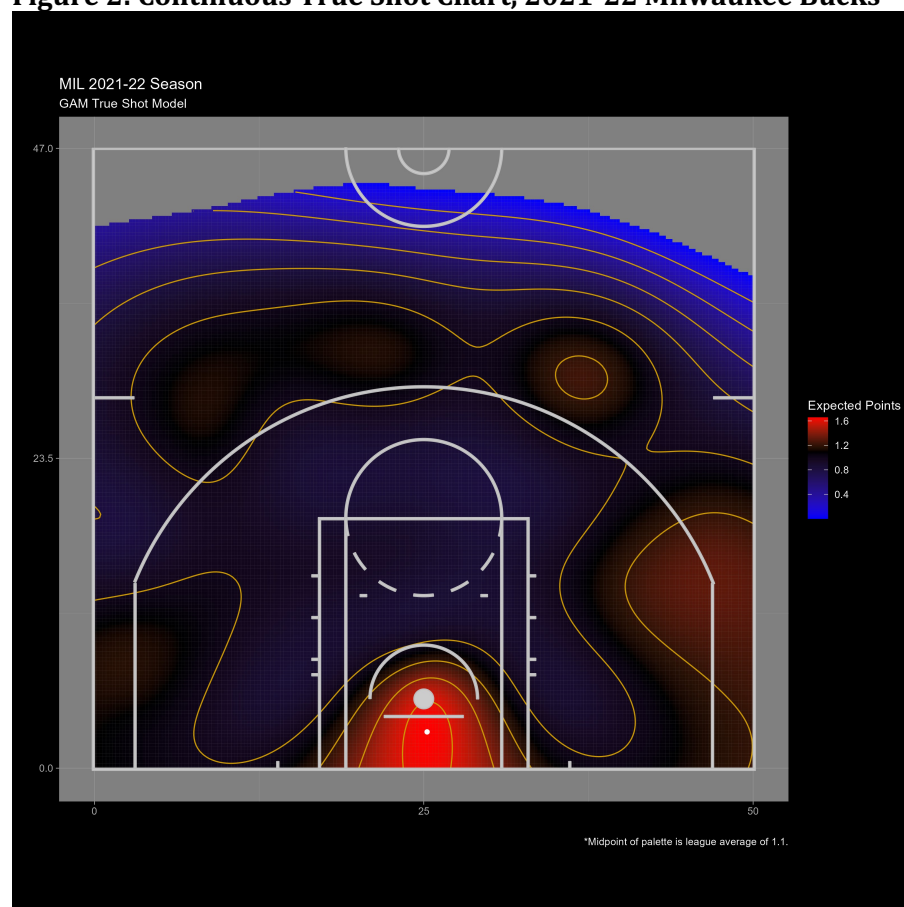
## 2. Shot Charts with Spatial Continuity and True Estimates of Point Yield

Shot charts help us understand trends in shot selection and the efficiency therefrom. To obtain a full picture of shot efficiency, however, we must consider all scoring pursuant to a given shot. Herein, we overcome limitations of previous shot chart types by developing a generalized additive model (GAM) continuous shot chart that accounts for all shot-pursuant free throw scoring. A GAM is a type of generalized linear model, where the outcome variable depends linearly upon smooth functions of the explanatory variables. In fitting a generalized additive model, these smooth functions are estimated. A GAM shot chart allows for the estimation of player or team shot efficiency as a continuous, three-dimensional surface, where the third dimension is represented not by the physical height of a surface at a particular (x,y) coordinate but via continuous color-coding. Hastie and Tibshirani (1987) first developed GAMs to serve as a flexible and interpretable family of non-linear models.

We call the GAM shot charts developed herein *Continuous True Shot Charts* and make them available on a custom, searchable project dashboard at <https://sportdataviz.syr.edu/TrueShotChart>. The project Github repository is available at: [https://github.com/Syracuse-University-Sport-Analytics/continous\\_shot\\_selection](https://github.com/Syracuse-University-Sport-Analytics/continous_shot_selection). Figure 2 shows a *Continuous True Shot Chart* for the 2021-22 Milwaukee Bucks. In a *Continuous True Shot Chart*, expected true point yield, from the field and line, for each possible shot location of the half-court, is estimated, and the estimate is color-coded into

the chart. Yellow ridge curves represent iso-yield curves or spatial level sets on yield such that shots along a given set generate equal expected true points.

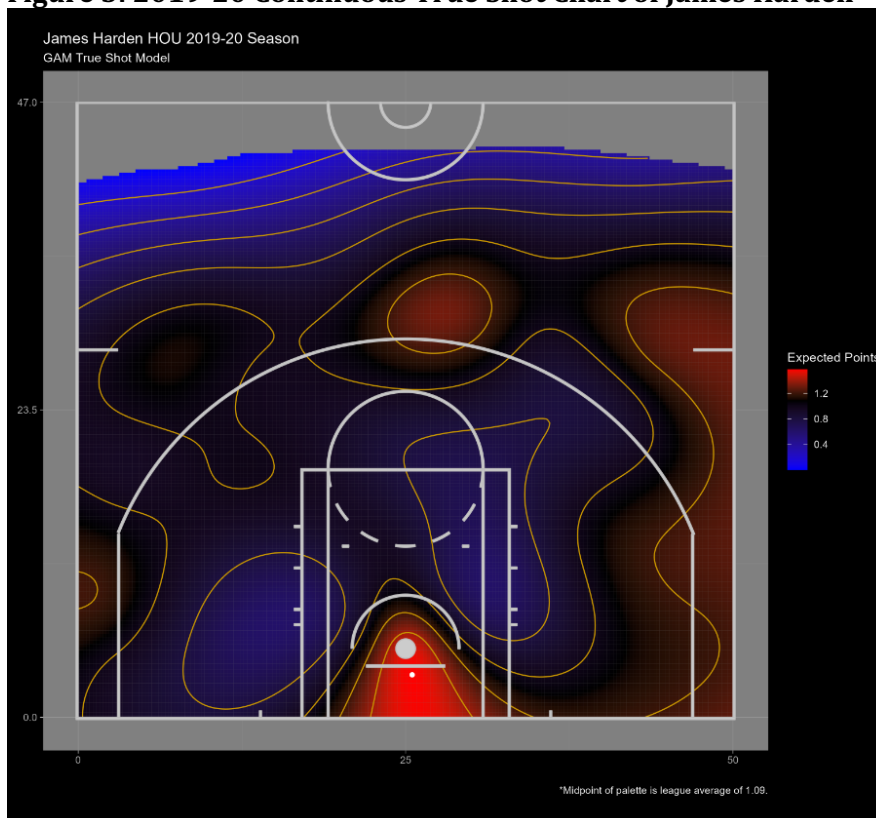
**Figure 2: Continuous True Shot Chart, 2021-22 Milwaukee Bucks**



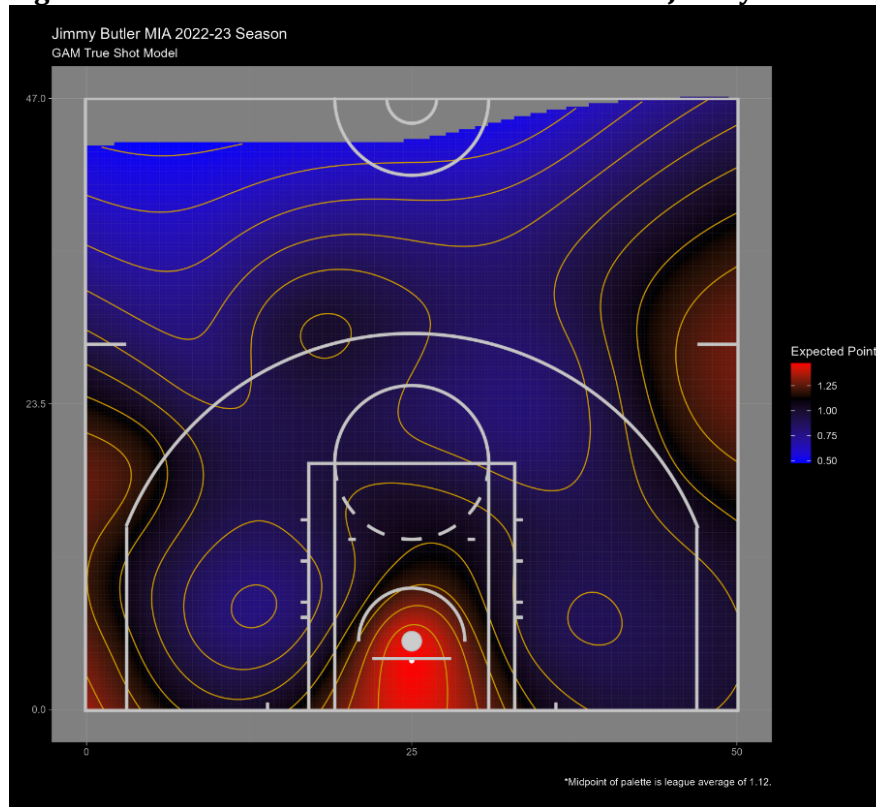
The present work represents the first example of a continuous (GAM-based) true shot chart, to the authors' best knowledge. When examining the 2021-22 Milwaukee Bucks' *Continuous True Shot Chart* of Figure 2, we observe some surprises and some expected results. When including shot-pursuant free throw scoring, we find the expected result that the Bucks scored proficiently at or near the rim, as well as from the wing three-point region extended and the left corner three-point area. However, we also find the surprising result that, for the Bucks of that season, long 2PAs were not nearly as low yielding as the conventional wisdom would suggest. The black far-midrange regions suggest that long 2PAs along the baseline extended were higher yielding than either close midrange shots in the paint or midrange shots off the elbows. Even more surprisingly, long 2PAs just inside the left three-point line break provided the Bucks with substantial yield (greater than 1.2 points per shot!). The 2021-22 Bucks obtained elite yield from a long 2PA region of the floor but middling yield for much of the near-midrange paint region. Further, the Bucks obtained elite scoring yield on 3PAs in some regions out to perhaps 26 feet, indicating that deep-threat 3PAs can have substantial scoring benefits in addition to the spacing benefits they afford an offense. However, any efficient scoring benefits of deep 3PAs fell off sharply beyond ~26 feet.

At its limit, a modern shot chart reduces a team's shooting regions down to 3PAs and 2PAs at or near the rim. When considering shot-pursuant free-throw scoring, we observe that these distilled areas are important but that there are potentially glaring exceptions to such reductionism. In Figure 3, we consider James Harden's 2019-20 shot chart. We chose this player-season for several reasons. This was Harden's last year in Houston, where he was the on-court leader of a revolution in shot chart reductionism, Daryl Morey having been the accompanying front office leader of said revolution. Further, Harden won his third consecutive scoring title that season. Lastly, Harden has an exceptionally high free throw rate such that we expect a great deal of additional information in his true shot chart. In that season, Harden's FT rate (FTA/FGA) was 0.557 or about 2.14 times the league average of 0.26. We then consider the 2022-23 shot chart of Jimmy Butler in Figure 4. We select Butler's shot chart because he also has an all-time FT rate, especially among fellow Small Forwards. In fact, the present authors cannot find a Small Forward with a higher seasonal FT rate than Butler. His FT rate has been as high as 0.709 for a season and was 0.625 in 2022-23. As such, we also expect Butler's true shot charts to look substantially different than his traditional chart.

**Figure 3: 2019-20 Continuous True Shot Chart of James Harden**



**Figure 4: 2022-23 Continuous True Shot Chart of Jimmy Butler**

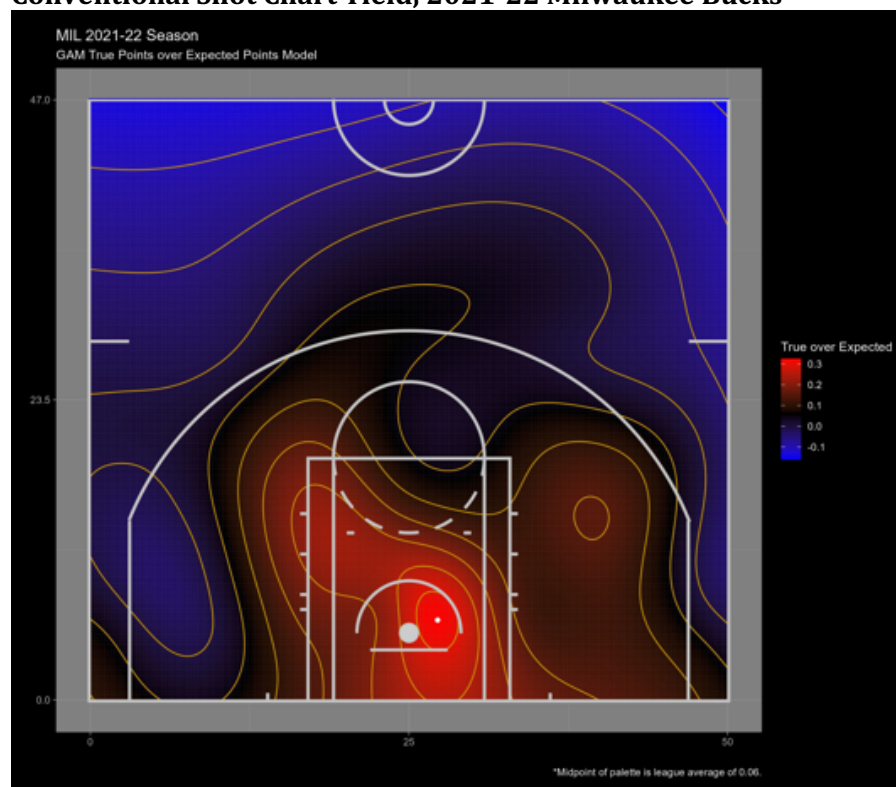


Figures 3-4 have rich commonalities and differences. Buoyed by prolific free throw activity, both charts feature the expected hot spots in parts of the three-point region, as well as at or near the rim. However, some aspects of these shot charts reinforce surprises from the previous chart of Figure 2. We observe areas of premium yield along the baseline extended long 2PA regions for Harden; these same areas are of moderate yield for Butler, rather than low. As in the case of Figure 2, we observe that premium long 2PA regions exist in the modern NBA shot chart when accounting for shot-pursuant FT scoring.

For the 2021-22 Bucks again, Figure 5 presents a differential shot chart that visualizes the difference, from each point in the half court, between the color coding, or yield, of the *Continuous True Shot Chart* and that of a conventional shot chart that does not include shot-pursuant FT scoring.



**Figure 5: Differential Shot Chart, Continuous True Shot Chart Yield minus Continuous Conventional Shot Chart Yield, 2021-22 Milwaukee Bucks**



We observe from Figure 5 that failing to include shot-pursuant FT scoring places a substantial, and typically downward, bias on the estimated yield of shots near the basket. In general, the differential is variable across the floor, suggesting a variable underlying shot position conditional FT rate. Thus, traditional shot charts can substantially distort spatial characterizations of scoring yield. These same results bear out for many other team seasons, as can be observed on our custom shot chart dashboard at <https://sportdataviz.syr.edu/TrueShotChart/>. Figure 5 also *explains* the unexpected result that, in a true shot chart, shots from the left baseline extended long 2PA region generated a high yield for the 2021-22 Bucks. Shots from this region generated a high made FT rate. It can be observed on the dashboard that James Harden's 2019-20 yield premium from this region also partly depended on a relatively high made FT rate for shots in that region.

Though largely reviled in the modern era of basketball analytics, long 2PAs in some cases generate a high true scoring yield. In fact, we observe from Figure 2 that long 2PAs from this region for the 2021-22 Bucks returned a similar point yield, on average, than if the player had stepped back to the adjacent three-point range! In this case, a comparison of Figures 2 and 5 suggests that the step-back 3PA garnered more points from the field but fewer points from the line. In consideration of the latter effect, step-backs may discourage defenders from fouling because a foul on a missed 3PA allows the offense 50% more points, in expectation, than fouling on a missed 2PA. Assuming league average FT proficiency, a missed 3PA that draws a shooting foul would garner an expected  $0.78 \cdot 3$  or 2.34 points, whereas a missed 2PA that draws a shooting foul would garner an expected  $0.78 \cdot 2$  or 1.56 points. This premium is certainly not commensurate with any three-point premium that exists in the absence of a shooting foul, as we will observe later in the paper. Therefore, fouling

on a 3PA disproportionately punishes the defense. This calculus likely explains at least part of the lower FT rate on 3PAs in Figure 5, as well as the lack of a sharp efficiency distinction between certain 3PAs and their neighboring long 2PAs. In the subsequent section, we aggregate *Continuous True Shot Charts* to determine whether results such as this suggest an erosion of the three-point premium in recent NBA seasons.

### 3. Assessing Shot Selection Efficiency from Continuous True Shot Charts

Another issue surrounding conventional shot chart analysis is the lack of a comprehensive measure by which to summarize a shot chart in terms of shot selection efficiency. The present work therefore further develops a shot selection measure called *Shot Selection Efficiency* that takes as input a player's or team's expected true points and expected proportional volume from a simulated grid of shots attempted somewhere on the half court and computes the player's or team's spatial Pearson correlation between expected proportional volume and expected true points, from the field and free throw line, across the half court.

This represents a measure of a player's or team's shot selection efficiency. The measure does not depend on overall shooting ability but, rather, on how proficient a player or team is in shooting from hot spots and avoiding cold spots on the floor, while taking into account shot location conditional free throw scoring rate. A separate measure of *exogenous shooting ability* is developed to estimate distinctly the average expected true points for a player (team) if they took a shot from every possible location on the shot chart with equal likelihood. As such, this latter variable estimates shooting ability but is exogenous of shot selection in that it weights each shot location equally. With this control measure, we are subsequently able to validate the statistical contribution or marginal effect of *Shot Selection Efficiency* to a basketball offense while controlling out the effect of *exogenous shooting ability*. We validate the *Shot Selection Efficiency* measure by specifying an XGBoost model, as well as linear, fixed effects regression models, at the team-season level that estimate Pythagorean expected win percentage conditional on a set of variables that survived a Variance Inflation Factor variable-paring model specification approach to limit multicollinearity. These include *exogenous shooting ability*, *defensive rating (DRtg)*, *own turnover rate (TOV%)*, *offensive rebounding rate (ORB%)*, *team fixed effects*, *Normalized Payroll*, and *Shot Selection Efficiency*. It stands to reason that these specific control variables survived VIF-paring. They encompass the non-shooting offensive four factors (i.e., the offensive four factors not related to shot selection efficiency or shooting ability) and the team's overall defensive effectiveness in that season. Table 1 summarizes the dependent and independent variables of the study.

**Table 1: Feature Summary Statistics (n=210)**

Characteristic	Median (SD) Range
Pyth.W.proportion	0.50 (0.14) 0.18 - 0.82
W.proportion	0.50 (0.14) 0.18 - 0.82
Net.Rating	0.40 (4.6) -10.5 - 11.6
Shot.selection.efficiency	0.27 (0.06) 0.01 - 0.44
Exogenous.shooting.ability	0.88 (0.05) 0.76 - 1.03
TOV%	12.60 (0.90) 9.90 - 14.90
ORB%	22.80 (2.21) 17.90 - 30.20
DRtg	111.2 (3.3) 102.9 - 120.0
Payroll (millions \$)	123.57 (22.32) 79.18 - 192.91
Normalized.Payroll	-0.02 (0.99) -2.93 - 2.59

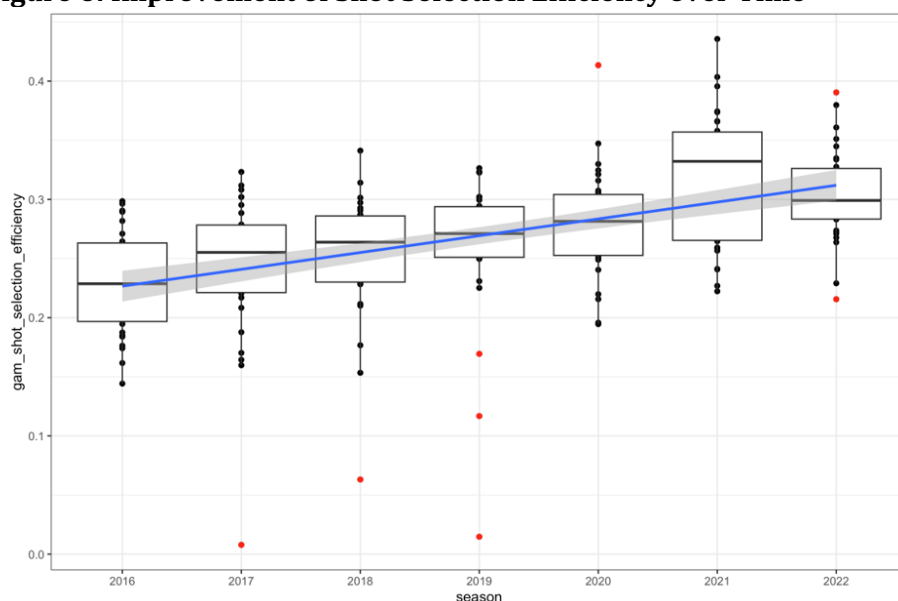
*Payroll normalized* represents a team's payroll in z-score terms, with respect to the distribution of NBA team payrolls in that season. *Net Rating* is a team's average scoring margin per 100 possessions, which approximates the average possessions per game in the 2022-23 NBA season. In 2022-23, the average points per NBA game was 114.7, and the average points per 100 NBA possessions was 114.8. Therefore, average possessions per game in that season was  $(114.7/114.8)$  or about 99.9. This value has been converging upward throughout the course of the NBA player-tracking movement, which has largely emphasized quick sets involving ball-movement, quickly-developing isolation plays, and pull-up 3PAs over slow sets involving foot-movement, slowly-developing isolation plays, and fewer pull-up 3PAs. Interestingly, rate stats for the NBA have been measured in *per 100 possession* terms for decades, and the NBA's analytic movement, which has greatly propelled the prominence of these rate stats, has helped to converge these *per 100 possession* stats to the magnitude of *per game* stats on league average. At least for the time being, this creates the advantage of less currency exchange in dealing with these two levels of statistics. In analyzing league-wide data for 2022-23, the choice of per game or per 100 possession stats is almost immaterial (material only at a margin of error less than 0.1%).

The median value for *Exogenous Shooting Ability* is 0.88, which matches the sample mean for this variable. This is interpreted as median true points per shot for a team that randomly selects shot locations, with equal likelihood, from across its shot chart. One might think of this team as one that plays "hot potato" on offense or one programmed to get the ball to any point in the half court with equal likelihood and then shoot. As mean and median are equal for this variable, such a team is expected to score 0.88 true points per shot compared to the observed sample average true points per shot of 1.13. We can meaningfully difference these two values to estimate that the net value of shot selection, on league average, in our sample is approximately a quarter of a point per shot taken.



All teams in the sample exhibit a positive *Shot Selection Efficiency* value. Recall that this variable represents the Pearson correlation between shot volume and true scoring from a grid of shot locations across the floor. The median value is 0.27, suggesting that the typical NBA team is mildly-to-moderately effective in identifying hot spots and avoiding cold spots in shot selection. The minimum observed value is 0.01, suggesting there was a team in a season (the 2017-18 Portland Trail Blazers) that had essentially no correlation between shot volume and true scoring in its shot selection across the floor! Luckily for that team, it had a high exogenous shooting ability such that their overall offense was effective. The maximum *Shot Selection Efficiency* was 0.44 and was set by the 2021-22 Boston Celtics, a team well-known for analytically-driven sets and shot selection. Another descriptive data summary question is whether the *Shot Selection Efficiency* has improved as we progress deeper into the analytic era of basketball. Figure 6 addresses this question with a time trend and embedded box plots for the variable by season.

**Figure 6: Improvement of Shot Selection Efficiency over Time**



In Figure 6, we observe a positive trend in expected team *Shot Selection Efficiency* over time. In 2016-17, the trend value was 0.23. By 2022-23, the trend value had risen to 0.32. With progressive advances in analytics, NBA offenses have improved in *Shot Selection Efficiency* over time. This trend result matches previous findings that offensive efficiency, and not defensive efficiency, is a leading indicator of analytic advances (see, e.g., Ehrlich and Sanders 2021). Having summarized the key model variables, a regression analysis will now estimate the respective marginal effects of *Shot Selection Efficiency*, *Exogenous Shooting Ability*, and other variables upon team-season win performance and net rating. Equations 1-3 provide the three primary model specifications.

Equation 1-3 for Win Production Models

$$E(\text{Pyth. W. proportion}) = \beta_0 + \beta_1 \cdot \text{Shot. Selection. Efficiency}_{ij} + \beta_2 \cdot \text{Exogenous. Shooting. Ability}_{ij} + \beta_3 \cdot \text{Vector. of. Controls}_{ij} \quad (1)$$

$$E(\text{W. proportion}) = \beta_0 + \beta_1 \cdot \text{Shot. Selection. Efficiency}_{ij} + \beta_2 \cdot \text{Exogenous. Shooting. Ability}_{ij} + \beta_3 \cdot \text{Vector. of. Controls}_{ij} \quad (2)$$

$$E(\text{Net. Rating}_{ij}) = \beta_0 + \beta_1 \cdot \text{Shot. Selection. Efficiency}_{ij} + \beta_2 \cdot \text{Exogenous. Shooting. Ability}_{ij} + \beta_3 \cdot \text{Vector. of. Controls}_{ij} \quad (3)$$

Model 1 estimates *Pythagorean Expected Win Proportion* for a team in a season as a function of the team's shot selection efficiency, exogenous shooting ability, and a vector of control variables in that season. Our Pythagorean exponent for the model is 14, updating from similar earlier estimates, e.g., by Morey (1993), such that a team's Pythagorean Expected Win Proportion in a season is given as

$\frac{ORTG^{14}}{ORTG^{14} + DRTG^{14}}$ . Model 2 features the same righthand side but with each team's (actual) *Win Proportion* as the dependent variable, while Model 3 substitutes *Net Rating* as the dependent variable. Results for the models are given In Table 2 as follows.

**Table 2: Shot Selection Efficiency Effect on Win Performance Models**

<i>Predictors</i>	<b>Pyth W Proportion</b>			<b>W Proportion</b>			<b>NetRtg</b>		
	<i>Estimates</i>	<i>std. Error</i>	<i>p</i>	<i>Estimates</i>	<i>std. Error</i>	<i>P</i>	<i>Estimates</i>	<i>std. Error</i>	<i>p</i>
(Intercept)	2.218	0.224	<b>&lt;0.001</b>	2.097	0.249	<b>&lt;0.001</b>	55.099	7.041	<b>&lt;0.001</b>
DRtg	-0.028	0.002	<b>&lt;0.001</b>	-0.027	0.002	<b>&lt;0.001</b>	-0.945	0.049	<b>&lt;0.001</b>
gam shot selection efficiency	0.400	0.086	<b>&lt;0.001</b>	0.334	0.095	<b>0.001</b>	14.149	2.691	<b>&lt;0.001</b>
exogenous shooting ability	1.307	0.107	<b>&lt;0.001</b>	1.393	0.119	<b>&lt;0.001</b>	49.073	3.355	<b>&lt;0.001</b>
off TOV%	-0.025	0.006	<b>&lt;0.001</b>	-0.021	0.007	<b>0.001</b>	-0.791	0.185	<b>&lt;0.001</b>
off ORB%	0.018	0.002	<b>&lt;0.001</b>	0.016	0.003	<b>&lt;0.001</b>	0.535	0.073	<b>&lt;0.001</b>
Observations	210			210			210		
R <sup>2</sup> / R <sup>2</sup> adjusted	0.719 / 0.712			0.672 / 0.664			0.757 / 0.75		

The regression data constitutes every NBA team-season from 2016-17 through 2022-23 (30 teams multiplied by 7 years). *Shot Selection Efficiency* has a positive and highly-significant conditional effect on *Pythagorean Expected Win Proportion*, *Win Proportion*, and *Net Rating*. If an NBA team in a

season can improve its weighted correlation between shot-volume and expected true points by 0.1 units, it gains an expected 0.04 units of win proportion or about 3.28 additional wins per 82-game regular season, according to the *Pythagorean Expected Win Proportion* model. In other words, a team's expected regular season win count increases by 1 for roughly every 0.03 units of increase in the spatial Pearson correlation between shot-volume and expected true points. To rephrase once more, a standard deviation increase in *Shot Selection Efficiency* yields the expectation of a 0.17 standard deviation increase in *Pythagorean Expected Win Proportion*, which suggests that variation in shot selection efficiency is a primary driver of winning in the NBA. From Model 3, we find that if an NBA team in a season can improve *Shot Selection Efficiency* by 0.1 units, it gains an expected 1.41 points in terms of average game-level score margin (*Net Rating*).

*Exogenous Shooting Ability* is also significant and substantial in explaining team performance, as are the other factor-type control variables. These results are both highly significant and substantial. Further, shot selection gains do not necessarily require a costly roster overhaul. In principle at least, shot selection efficiency is something that can be taught to players. We next examine whether subsequent regressions will determine whether shot selection efficiency is a “moneyball,” or inefficiently-priced source of wins. The three models feature strong explanatory power ( $R^2$  between 0.672 and 0.757).

## 4. Is Shot Selection Efficiency Money(ball)?

Another model was estimated to consider the payroll-conditional effect of *Shot Selection Efficiency* on team success by adding the features *Normalized Payroll* and *Normalized Payroll Squared* to Model 1. As stated, the *Normalized Payroll* feature is the raw payroll standardized (center and scale) by season, which is calculated by subtracting the mean of payroll and then dividing by the standard deviation. We also included a *Normalized Payroll Squared* feature to allow for a possible quadratic payroll effect (due, e.g., to the possibility of some large-market teams being managerially inefficient). Table 3 models represent payroll-augmented linear and quadratic versions of Model 1.

**Table 3: Moneyball Shot Selection Efficiency Opportunity**

<i>Predictors</i>	Pyth W Proportion			Pyth W Proportion		
	<i>Estimates</i>	<i>std. Error</i>	<i>P</i>	<i>Estimates</i>	<i>std. Error</i>	<i>P</i>
(Intercept)	2.218	0.224	<b>&lt;0.001</b>	2.143	0.221	<b>&lt;0.001</b>
DRtg	-0.028	0.002	<b>&lt;0.001</b>	-0.026	0.002	<b>&lt;0.001</b>
gam shot selection efficiency	0.400	0.086	<b>&lt;0.001</b>	0.379	0.084	<b>&lt;0.001</b>
exogenous shooting ability	1.307	0.107	<b>&lt;0.001</b>	1.195	0.111	<b>&lt;0.001</b>
off TOV%	-0.025	0.006	<b>&lt;0.001</b>	-0.023	0.006	<b>&lt;0.001</b>
off ORB%	0.018	0.002	<b>&lt;0.001</b>	0.018	0.002	<b>&lt;0.001</b>
normalized payroll				0.018	0.006	<b>0.002</b>
normalized payroll sq.				-0.003	0.003	0.336
Observations	210			210		
R <sup>2</sup> / R <sup>2</sup> adjusted	0.719 / 0.712			0.733 / 0.724		

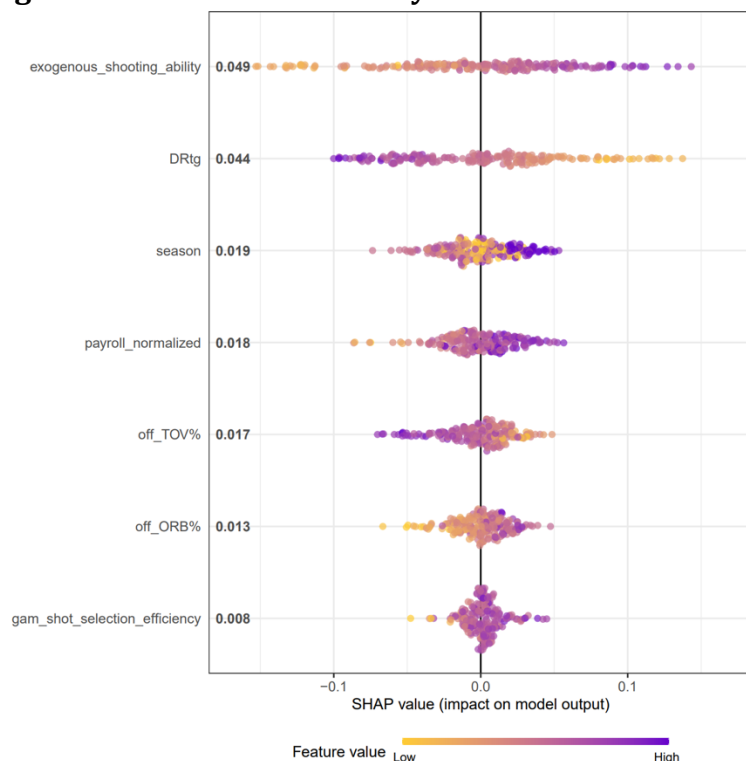
Payroll is significantly productive of expected wins in both models. Further, the quadratic term on *Normalized Payroll* is negative but insignificant in Model 2. The *in-sample* relationship between payroll and expected wins is quadratic (inverted-U), but the coefficient on the quadratic payroll term is insignificant. Even conditional on payroll, *Shot Selection Efficiency* is positive and significant, indicating that the win productivity of *Shot Selection Efficiency* is not explained by variation in team-season payroll. This may suggest either that players have different observed levels of *Shot Selection Efficiency* but that the player labor market is not valuing higher levels of the attribute or that team coaching and analytic staffs (i.e., team culture and values) are largely responsible for variation in team *Shot Selection Efficiency*. In either case, *Shot Selection Efficiency* is a “moneyball” or supra-payroll source of wins in the NBA. The magnitude of the coefficient for *Shot Selection Efficiency* remains essentially the same when we control for team-season payroll. This suggests that the NBA player labor market is essentially not pricing in any variation in *Shot Selection Efficiency* observed at the player level. Rather, this variation is driven by team coaching/analytics.

## 5. Robustness Check of Shot Selection Efficiency using XGBoost

To robustness check our linear, fixed effect model results, we fit an XGBoost model (Liu and Lust, 2020) to understand how *Shot Selection Efficiency*, along with other covariates, relate to *Pythagorean Expected Win Proportion*. The covariates of interest include *Normalized Payroll*, *DRtg*, *ORB%*, *TOV%*, *Exogenous Shooting Ability*, and *Season*. We partitioned 90% of the data into a training set and 10% into a test set. The out-of-sample mean absolute error was 0.052 and the out

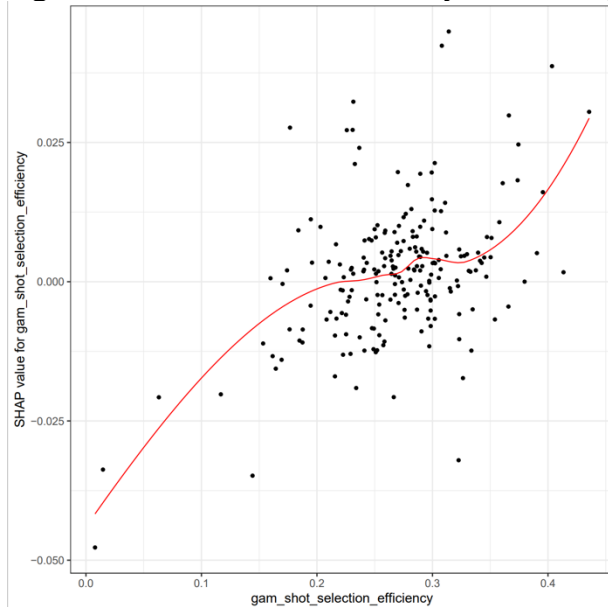
of sample R-squared was 0.625, comparable to that observed in the linear, fixed effects models. For the full model, the mean absolute error was 0.0004, and the R-squared was 0.999989. Also, Shapley Additive exPlanations (SHAP) (Lundberg and Lee, 2017) values were calculated. Figure 7 shows a beeswarm summary plot, which demonstrates the estimated effect of each parameter on expected wins.

**Figure 7: Beeswarm Summary Plot**

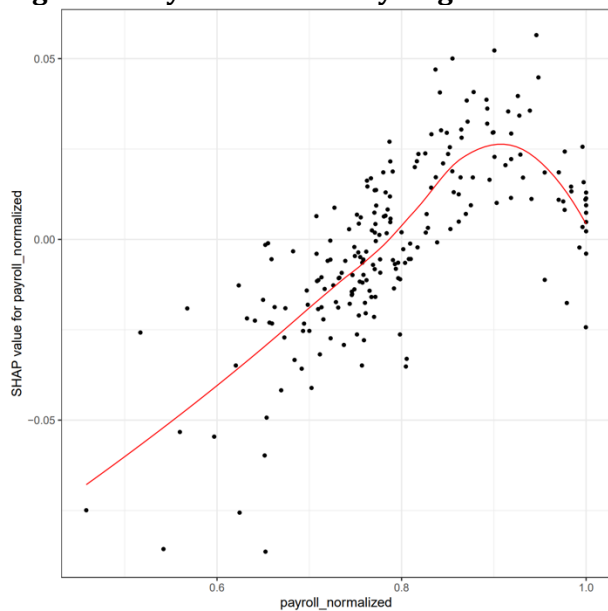


*Shot Selection Efficiency* has a positive estimated Shapley Value such that the XGBoost results are consistent with those of the linear, fixed effects models. Two SHAP dependency plots were also rendered, where each shows estimated marginal effect that each feature has on Pythagorean Expected Win Proportion. Figure 8 demonstrates the monotonically positive impact that *Shot Selection Efficiency* has on *Pythagorean Expected Win Proportion*. This confirms that shot selection efficiency is indeed important to team performance. For teams with a *Shot Selection Efficiency* greater than 0.23, shot selection positively contributed toward winning, according to Figure 8.

**Figure 8: Shot Selection Efficiency Effect on Pythagorean Wins**



**Figure 9: Payroll Effect on Pythagorean Wins**



## 6. Does an NBA 3PA Premium Remain according to Aggregations of True Shot Charts?

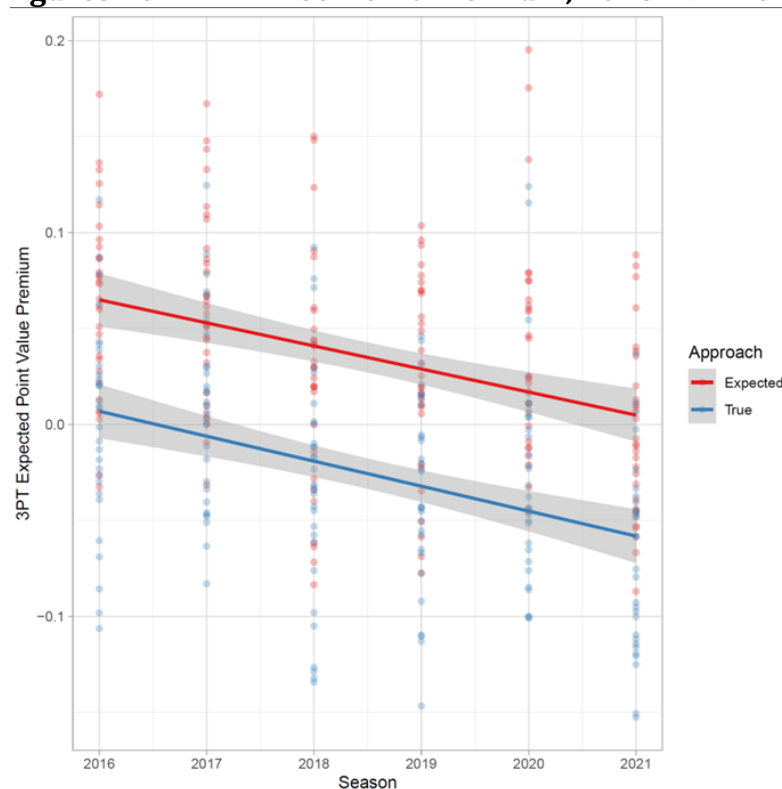
By accounting for FT scoring, *Continuous True Shot Charts* can comprehensively address many fundamental questions in basketball, such as: *Does a three-point premium still exist in the NBA given the escalation in three-point attempt volume?* See, for example, Ehrlich and Sanders (2021), who develop the first discrete, true shot chart to find evidence that the three-point premium eroded to



some degree during the 2010s. While aggregations of *conventional* shot charts suggest the three-point yield premium remains intact for the NBA, what happens when we consider shot-pursuant FT scoring for recent seasons?

To consider this question, we aggregate both Continuous True Shot Charts and traditional shot charts for all NBA shots taken in each season from 2016-17 through 2021-22. Doing so allows us to estimate the three-point premium, with and without shot-pursuant free throw scoring in each season. Figure 10 reports those results, where the red trend line is estimated from aggregations of traditional shot charts and therefore reports the estimated difference between 3PA and 2PA yield when not considering shot-pursuant FTAs. The blue line is estimated from aggregations of Continuous True Shot Charts and therefore reports the estimated difference between 3PA and 2PA yield when considering shot-pursuant FTAs.

**Figures 10: NBA Three-Point Premium, 2016-17 – 2021-22: A Tale of Two Shot Charts**



Observing the red trend line, we find a significant and fairly substantial leaguewide three-point *premium* that existed from 2016-17 through 2020-21 when considering aggregations of *traditional* shot charts. That is, NBA players scored significantly more proficiently *from the field* on three-point attempts than on two-point attempts from 2016-17 through 2020-21. In magnitude, this premium ranged between 0.025 and 0.064 points per three-point shot over these years. In 2021-22, there was no significant leaguewide three-point premium or dispremium when considering aggregations of *traditional* shot charts. However, when considering aggregations of *true shot charts*, we find a significant and deepening three-point *dispremium* at the league level since the 2018-19 season! In 2021-22, this dispremium swelled to -0.066 points per three-point shot or roughly a point less for each 15 three-point shots. Despite these findings, we observe that leaguewide three-point shot volume has continued to rise dramatically. It appears as though the typical team continued to follow *perceived* three-point premia from aggregations of traditional shot charts rather than

significant disparity from aggregations of *true shot charts*. In the case of 3PA volume, there is evidence that NBA teams overshot their market, assisted by a mis-specified analytic tool in the form of traditional shot charts.

## 7. Conclusion

In this work, we have developed a set of novel GAM-estimated, Continuous True Shot Charts for NBA teams and players and have demonstrated that they correct distortions in scoring yield estimates found in traditional shot charts. They do so by including shot-pursuant FT scoring into the color coding (scoring yield estimate) of the chart at all locations. We further construct a measure of team shot selection efficiency as a means to summarize team and player shot charts in terms of efficiency. This measure represents the Pearson correlation between expected proportional volume and expected true points, from the field and free throw line, across the half court. Through linear, polynomial, and XGBoost modeling, we find that shot selection efficiency contributes to team expected win proportion significantly and substantially. Further, shot selection efficiency is found to be a “moneyball” or supra-payroll source of wins in the NBA. Lastly, we use Continuous True Shot Charts to find that there has been a widening NBA three-point disparity since 2018-19. This disparity result is not available if one aggregates from traditional shot charts, which do not consider shot-pursuant FT scoring. Continuous True Shot Charts are both novel and meaningful in basketball league settings.

## 8. References

- Ehrlich, J. A., Ghimire, S., Sadler, T. R., & Sanders, S. D. (2022). Policy and Policy Response on the Court: A Theoretical and Empirical Examination of the Three-Point Line Extension in Basketball. *Journal of Sports Economics*, 15270025221111790.
- Fichman, M., & O'Brien, J. R. (2019). Optimal shot selection strategies for the NBA. *Journal of Quantitative Analysis in Sports*, 15(3), 203-211.
- Franks, A., Miller, A., Bornn, L., & Goldsberry, K. (2015, February). Counterpoints: Advanced defensive metrics for NBA basketball. In *9th Annual MIT Sloan Sports Analytics Conference*, Boston, MA.
- Goldman, M., & Rao, J. M. (2011, March). Allocative and dynamic efficiency in NBA decision making. In *In Proceedings of the MIT Sloan Sports Analytics Conference* (pp. 4-5).
- Goldsberry, K. (2019). *Sprawlball: A visual tour of the new era of the NBA*. Mariner Books.
- Hastie, T., & Tibshirani, R. (1987). Generalized additive models: some applications. *Journal of the American Statistical Association*, 82(398), 371-386.
- Jiao, J., Hu, G., & Yan, J. (2021). A Bayesian marked spatial point processes model for basketball shot chart. *Journal of Quantitative Analysis in Sports*, 17(2), 77-90.
- Jieying, J., Guanyu, H., & Jun, Y. (2021). A Bayesian marked spatial point processes model for basketball shot chart. *Journal of Quantitative Analysis in Sports*, 17(2), 77-90.
- Lundberg, S. M., & Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions*. <https://arxiv.org/abs/1705.07874>
- Massey, C., & Thaler, R. H. (2013). The loser's curse: Decision making and market efficiency in the National Football League draft. *Management Science*, 59(7), 1479-1495.
- Morey, D. 1993. “STATS Basketball Scoreboard.” Pp. 1-288 in *STATS Basketball Scoreboard*, edited by J. Dewan and D. Sminda. New York: STATS, Inc.

- Narayan, S. (2019). *Applications of machine learning: basketball strategy* (Doctoral dissertation, Massachusetts Institute of Technology).
- NBA. (2022). Example Shot Chart. Available Online at:  
<https://www.nba.com/stats/events/?flag=3&CFID=33&CFPARAMS=2021-22&PlayerID=201939&ContextMeasure=FGA&Season=2021-22&section=player&sct=hex>
- NPR (2020). The Science of Hoops. Available online at:  
<https://www.npr.org/transcripts/911898347>
- Otto, C. A., & Volpin, P. F. (2018). Marking to market and inefficient investment decisions. *Management Science*, 64(8), 3756-3771.
- Papalexakis, E., & Pelechrinis, K. (2018, October). thoops: A multi-aspect analytical framework for spatio-temporal basketball data. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management* (pp. 2223-2232).
- Piette, J., Anand, S., & Zhang, K. (2010). Scoring and shooting abilities of NBA players. *Journal of Quantitative Analysis in Sports*, 6(1).
- Reich, B. J., Hodges, J. S., Carlin, B. P., & Reich, A. M. (2006). A spatial analysis of basketball shot chart data. *The American Statistician*, 60(1), 3-12.
- Romer, D. (2006). Do firms maximize? Evidence from professional football. *Journal of Political Economy*, 114(2), 340-365.
- Sandholtz, N., Mortensen, J., Bornn, L. (2019) Chuckers: Measuring Lineup Shot Distribution Optimality Using Spatial Allocative Efficiency Models Sloan Sports Analytics Conference 2019.
- Santos-Fernandez, E., Denti, F., Mengersen, K., & Mira, A. (2022). The role of intrinsic dimension in high-resolution player tracking data—Insights in basketball. *The Annals of Applied Statistics*, 16(1), 326-348.
- Skinner, B., & Goldman, M. (2015). Optimal strategy in basketball. *arXiv preprint arXiv:1512.05652*.
- Winston, W. L., Nestler, S., & Pelechrinis, K. (2022). 37. SportVU, Second Spectrum, and the Spatial Basketball Data Revolution. In *Mathletics* (pp. 321-340). Princeton University Press.
- Yang Liu and Allan Just (2020). SHAPforxgboost: SHAP Plots for 'XGBoost'. R package version 1.7.5.1. <https://github.com/liuyanguu/SHAPforxgboost/>