STAT 202B John Baierl
Final Paper

Analysis of NBA Team-Level Statistics and Performance

Abstract:

Despite ever-increasing access to a wide array of metrics in sports statistics, one of the central challenges in assessing the relevant factors for team success remains the limited scope of each individual measurement in the context of a dynamic and multifaceted game. In this paper, we analyze team-level statistics gathered during the 2021-22 NBA season to identify the primary dimensions in which teams vary in their performance, and the extent to which those measurements are meaningful predictors of regular season success. We find that in general, NBA teams vary more significantly in their offensive output than on the defensive end, with offensive performance being the primary driver of success. Moreover, traditional defensive counting statistics like steals and blocks fail to accurately capture performance on that end of the floor.

I. Introduction

Nearly all major sports have undergone some version of an analytics revolution over the past several decades, with improved access to data transforming the way that each game is played, managed, and evaluated. While this has often involved subtle shifts in decision-making and talent evaluation, the transformation of NBA basketball is a particularly illustrative instance of this trend, with analytics completely transforming the basic strategic premises of both offensive and defensive gameplay. Many of these changes are apparent just by the eye test and can be verified by singular traditional statistics, such as an increasing pace of play and the now well-documented supplanting of long-range twos for three-point shots (*Fig. 1*). The result is an on-court product that is often unrecognizable from the slow-moving, isolation-heavy playstyle of the late 90s and early 00s.

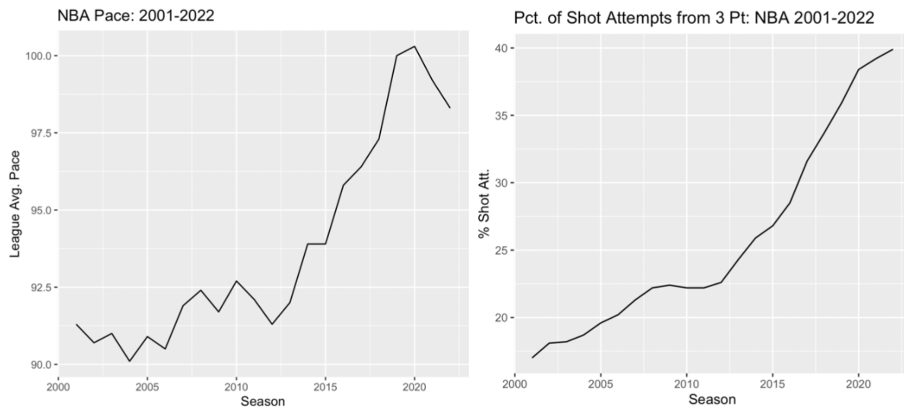

Fig. 1

However, subtler shifts are more difficult to detect. A frequent hole in the scope of NBA analysis is the assumed independence of a team's offensive and defensive behavior on the court. Is a team that playing a fast-paced offense sacrificing defense as a result? Is the trend toward more three-point shooting de-emphasizing the importance of defense in general? Which traditional statistics are becoming more or less informative in predicting winning in this changing landscape? These are central questions for long-term team planning and talent evaluation, motivating a deeper data-analytic inquiry into underlying trends as the modern NBA continues to evolve.

II. Data Overview

The primary dataset we will consider was taken from the 2020-21 NBA season, obtained from Basketball Reference (see References). The available data combines both traditional cumulative box score metrics such as points, rebounds, and assists for all 30 teams, as well as more modern statistics adjusting for pace and shot location. These include three-point attempt rate and true shooting percentage, among others.

The most widely-used summary statistics used to assess overall team performance are "offensive rating" (ORtg) and "defensive rating" (DRtg). Offensive rating computes the number of points scored per 100 possessions, while defensive rating computes the number of points allowed per 100 possessions, effectively quantifying a team's efficiency by adjusting for differences in pace of play. Taken together, these provide a general picture of where a team's performance sits relative to league average on both ends of the court. These values are plotted in *Fig. 2* for all 30 teams in the 2020-21 NBA season, with their final number of wins indicated by color (the 2020-21 season consisted of only 72 games, as opposed to the usual 82). Note that the defensive rating axis is reversed in this plot to aid in visual intuition, since teams with a higher numerical value in DRtg performed worse defensively.

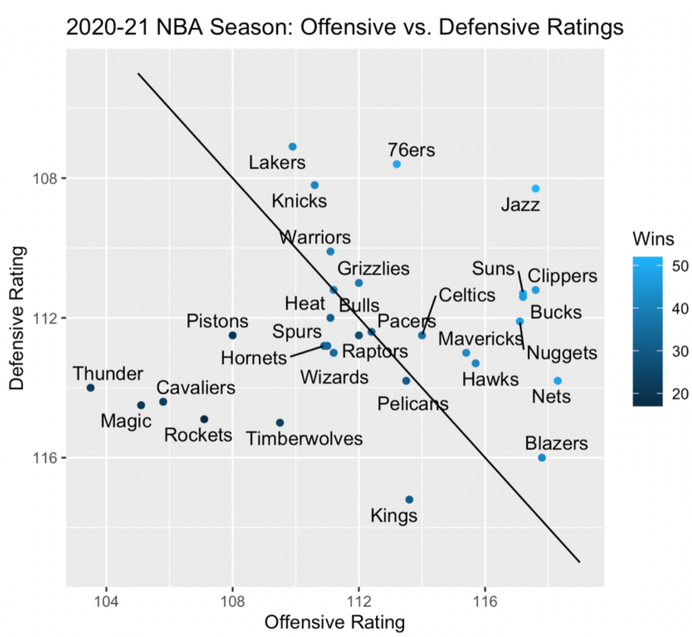

Fig. 2

The black diagonal line indicates a constant *net rating* ($Net\ Rtg = ORtg - DRtg$), passing through the point of league average ($ORtg = DRtg = 112.3$). Intuitively, points along this line represent teams of roughly league average overall quality, but with varying balances of offense and defense. So, net positive teams will fall in the upper right portion of the graph, and net negative teams in the lower left portion.

Several trends are apparent here. First, there is a greater level variability in the quality of NBA offenses that there is in the quality of NBA defenses based on the spread along each axis. Indeed, the standard deviation of offensive ratings is nearly double that of defensive ratings ($sd(ORtg) = 4.00$, $sd(DRtg) = 2.38$). Secondly, only the top five teams in overall wins (Jazz: 52, Suns: 51, 76ers: 49, Nets: 48, Nuggets: 47), only one possessed a defense more highly rated than their offense (76ers). Indeed, the teams with the top three defenses finished 10th (Lakers), 3rd (76ers), and 11th (Knicks) in overall wins. In comparison, the teams with the top three offenses finished 4th (Nets), 9th (Blazers), and 1st (Jazz) in overall wins. The Blazers notably did so while maintaining the second worst defense in the league.

In general, this provides some indication that offensive and defensive quality are not equally important in determining a team's success in the regular season. Moreover, there is no obvious indication from the plot above of offensive and defensive performance being a zero-sum game, where teams choose to emphasize one or the other through their lineup construction, with variability emerging primarily from differences in team philosophy.

Basketball statistics encompass a wide variety of metrics that attempt to measure the same concept in different ways. This is an inherent challenge when attempting to quantify a something as general as offensive or defensive "effectiveness." Each statistic paints a slightly different but overlapping picture that, when considered together, improve our basis for performance assessment. However, a consequence of this is that we should be cautious of a degree of multicollinearity among those statistics. A correlation matrix is given below (*Fig. 3*) among a set of 14 team-level statistics taken over the 2021-22 NBA season for all 30 teams. A guide for the variable abbreviations is provided in *Fig. A* in the appendix.

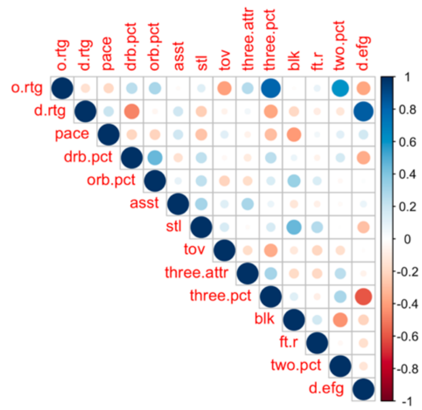

Fig. 3: Correlation Matrix

While this does not reveal a tremendously high degree of obvious multicollinearity, it highlights some of the unexpected associations between these variables. As we might expect, offensive rating is highly correlated with three-point shooting percentage ($\rho = 0.800$), and to a lesser degree, two-point shooting percentage ($\rho = 0.610$). But less immediately expected is the negative correlation between three-point shooting percentage and defensive effective field goal percentage ($\rho = -0.599$).

This gives a window into the challenges of determining the causal structure behind these relationships. It is possible that individual players that prioritize three-point shooting tend to be worse defenders. Or that lineups that prioritize generating open three-point shots tend to be smaller, leading to less effective defense.

III. Analysis and Discussion

We will build on this analysis in two ways, first by performing a principal component analysis on select performance metrics, then by modelling their relationship with the output of total win count through a ridge regression.

a) *PCA*

Given the high level of variation in playstyles, analyzing the trends and associations within the wide variety of metrics available to NBA teams is worthwhile to identify broader tendencies throughout the league. 14 different variables were selected to measure different elements of team performance on both offense and defense. Variables were selected to encompass as many elements of team performance as possible. Obvious redundancies were avoided as much as possible, such as eliminating all-encompassing metrics like true shooting percentage in favor of multiple measures to provide higher resolution, like free throw rate, three-point percentage, and two-point percentage. In general, variables were selected to adjust for differences in pace and shot distribution, such as selecting *assists per 100 possessions* was cover simply *assists per game*. The full list of included variables, as well as their abbreviations in all plots is provided in *Fig. A* in the appendix. All variables were standardized before carrying out the PCA.

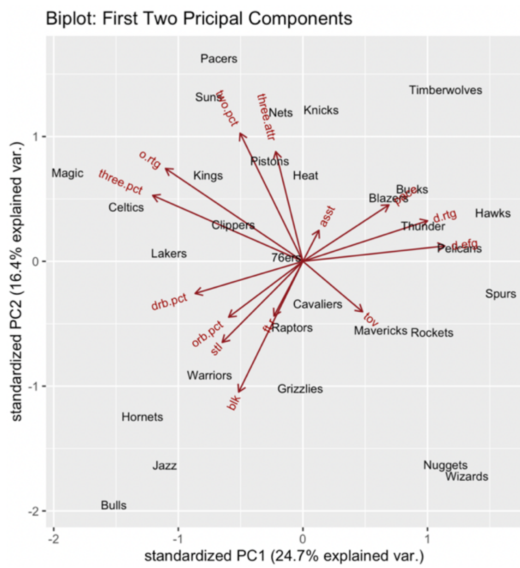

Fig. 3

The first two principal components are displayed on the biplot in *Fig. 3*. Overall, the first principal component accounts for only 24.7% of the total variation in the data. While this is relatively low, it is important to keep in mind that the variables considered encapsulate a variety of different elements of the game on both the offensive and defensive ends of the floor. As such, we should not expect to see especially high values here. Indeed, having over 41% of the total variation accounted for by the first two principal component is surprisingly high given the variety of playstyles and skillsets throughout the league.

The first principal component is most strongly correlated with three-point shooting percentage (–0.430), defensive effective field goal percentage (0.406), and offensive rating (–0.394). This underlines the importance of three-point shooting in the modern NBA, suggesting that it is among the biggest sources of variability between teams. The biplot above shows how closely aligned three-point shooting percentage and offensive rating are as well, lending some justification to its increased emphasis in recent years. However, note that the alignment between three-point shooting percentage and three-point attempt rate is not particularly strong ($\rho =$

0.310, from correlation matrix), indicating that teams three-point attempts are less correlated with their shooting aptitude than we might expect.

The second principal component is most strongly correlated with block rate (–0.460), two-point shooting percentage (0.450), and three-point attempt rate (0.386). Two-point shot attempts range from extremely high percentage shots (dunks and layups) to challenging mid-range looks from 16-23 ft, suggesting that shot selection likely plays a role in this variability. Block rate is highly personnel-dependent, with elite defensive centers tending to dominate blocking statistics per Basketball Reference. However, the biplot above shows little correlation between block rate and overall defensive rating, suggesting that shot blocking by itself is not strongly suggestive of a high-quality defense.

a) Ridge regression

While the previous analyses seek out underlying structure within the dataset, we have yet to establish any clear connection between these predictors and the desired outcome of winning games. To do this, we perform a ridge regression, opting against a principal component regression due to the relatively small amount of variability explained by the first few principal components. 10-fold cross validation was performed to tune the ridge parameter, adopting a value of $\lambda = 10.35$ using the one standard error heuristic. Given the differing scales of the 14 selected variables, we standardize each before regressing. A plot of the resulting mean squared error at each parameter value is provided in the appendix (*Fig. B*). The resulting regression coefficients are reported below. The coefficient of determination for this fitted model is $R^2 = 0.831$, suggesting reasonably strong predictive power.

(Int)	***o.rtg***	***d.rtg***	***pace***	***drb.pct***	***orb.pct***	***asst***	***stl***	***tov***	***three. attr***
36.0	2.29	–1.29	–0.558	0.605	0.276	–0.477	0.208	–0.478	–0.081
three.pct	**blk**	**ft.r**	**two.pct**	**d.efg**					
1.83	–0.255	0.234	1.45	–1.56					

Table 1: Ridge regression coefficients (see appendix for variable key)

The added emphasis on three-point shooting in recent years is clearly justified by these regression results. After offensive rating, improving three-point shooting results in the greatest increase in predicted win total among the offensive metrics considered. However, it is noteworthy that two-point shooting percentage is not far behind. In fact, a percentage point increase in two-point shooting percentage is roughly comparable with a percentage point decrease in opponent effective field goal percentage in terms of predicted wins in this model. This gives some indication that two-point shooting percentage remains a worthwhile goal to pursue, despite the increasing prevalence of long-distance shooting in the NBA.

Taken with the PCA analysis above, this corroborates the observation that, while defense remains a meaningful driver of winning NBA games, offensive statistics are responsible for a higher degree of variation between teams. These regression results suggest that defensive impact is secondary to offensive performance in producing regular season wins. Moreover, traditional defensive counting statistics (blocks, steals) are relatively poor predictors of regular season success in general, and do not correlate strongly with the overall quality of a defense.

IV. Conclusions and Further Questions

These analyses support that claim that offensive and defensive performance are not on equal footing in the modern NBA. Indeed, the Portland Trailblazers' performance in the 2021-

22 season, finishing 9th in wins with the 2nd ranked offense but 28th ranked defense, may not be so atypical, albeit an extreme example. Moreover, the impact of a defense is poorly accounted for by simply measuring blocks and steals, the two traditionally cited defensive box score statistics. Future analysis of NBA defensive impact should focus more on metrics like opponent shot selection and distance to nearest defender—factors that are more likely to consistently impact opponent field goal percentage, which our analysis shows is predictive of regular season success. Additionally, non-parametric models that do not impose such strong linearity assumptions may be worth exploring here. While improving defensive efficiency may be a viable route for improving a poor-performing team, there may be a ceiling beyond which elite offense is necessary to continue seeing improvement.

It should also be noted that this analysis was limited to regular season results, with playoff performance completely absent here. Among the four playoff semi-finalists (Bucks, Suns, Hawks, Clippers) all possessed top-8 offensive ratings but with defenses ranging from 3rd (Suns) to 27th (Hawks). Exploring the extent to which regular season metrics translate to postseason performance is an open question that warrants further study, but there are early indications here that these trends carry over. Expanding this data to span multiple seasons and separating regular season from postseason performance would help to solidify many of these inferences.

References:

Basketball Reference: full data set (https://www.basketball-reference.com/leagues/NBA_2022.html)

Basketball Reference: leaders in shot blocks per game (https://www.basketball-reference.com/leagues/NBA_2022_per_game.html#per_game_stats::blk_per_g)

Appendix:

Statistic:	**Biplot Abbreviation:**
Offense rating (points scored per 100 poss.)	o.rtg
Defense rating (points allowed per 100 poss.)	d.rtg
Pace (possessions per 48 min)	pace
Defensive Rebound Percentage (% opponents shot attempts rebounded)	drb.pct
Offensive Rebound Percentage (% own shot attempts rebounded)	orb.pct
Assist rate (assists per 100 possessions)	asst
Steal rate (steals per 100 possessions)	stl
Turnover rate (turnovers per 100 possessions)	tov
3-point attempt rate (% of shot attempts from 3 pt)	three.attr
3-point percentage (% of 3 pt. shot attempts made)	three.pct
2-point percentage (% of 2 pt. shot attempts made)	two.pct
Block rate (blocks per 100 possessions)	blk
Free throw rate (number of free throws per shot attempt)	ft.r
Defensive effective field goal percentage (weighted FG% allowed on def.)	d.efg

Table. A

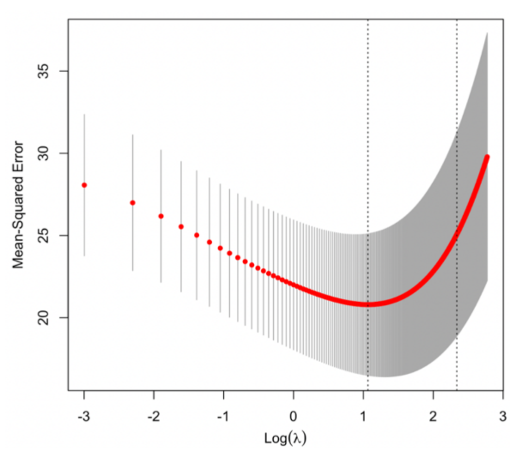

Fig. B: 10-Fold Cross Validation Tuning Ridge Parameter