Scoring and Shooting Abilities of NBA Players

James Piette, University of Pennsylvania
Sathyanarayan Anand, University of Pennsylvania
Kai Zhang, University of Pennsylvania

Recommended Citation:
Available at: http://www.bepress.com/jqas/vol6/iss1/1
DOI: 10.2202/1559-0410.1194

©2010 American Statistical Association. All rights reserved.
Scoring and Shooting Abilities of NBA Players

James Piette, Sathyanarayan Anand, and Kai Zhang

Abstract

We propose two new measures for evaluating offensive ability of NBA players, using one-dimensional shooting data from three seasons beginning with the 2004-05 season. These measures improve upon currently employed shooting statistics by accounting for the varying shooting patterns of players over different distances from the basket. This variance also provides us with an intuitive metric for clustering players, wherein performance of players is calculated and compared to his cluster center as a baseline. To further improve the accuracy of our measures, we develop our own variation of smoothing and shrinkage, reducing any small sample biases and abnormalities.

The first measure, SCAB or, Scoring Ability Above Baseline, measures a player's ability to score as a function of time on court. The second metric, SHTAB or Shooting Ability, calculates a player's propensity to score on a per-shot basis. Our results show that a combination of SCAB and SHTAB can be used to separate out players based on their offensive game. We observe that players who are highly ranked according to our measures are regularly considered as top performers on offense by experts, with the notable exception of LeBron James; the same claim holds for the offensive dregs. We suggest possible explanations for our findings and explore possibilities of future work with regard to player defense.

KEYWORDS: basketball, NBA, shooting, offense, field goal percentage

Author Notes: We would like to thank Shane Jensen and Dylan Small for reviewing our work and providing us with helpful commentary.
1 Introduction

General managers in any sport are constantly searching for better methods of player evaluation and basketball is no exception. There are many aspects of an NBA player that a general manager must consider when performing their analysis: offensive and defensive ability, makeup, personality, etc. Many of these skills, especially those pertaining to offense and defense, can be quantified using statistical measures. One of the most common measures in basketball pertains to offensive ability: field goal percentage (FG%)\(^1\) or some variant of it.

FG% is a reasonable estimate of a basketball player’s ability to shoot the ball, but it is not unbiased. To illustrate this point, let us consider the following example. Consider one player who concentrates on taking shots in the paint, or within about ten feet from the basketball hoop. Also consider another player that focuses on shooting from the three point line, which is about 23 feet to the basket.

Under the assumptions that it is harder to shoot the farther away you move from the basket\(^2\) and that the two players have the same shooting ability, we would expect that the player shooting in the paint will have a higher FG% than the three point shooter. In fact, we observe this phenomenon in the NBA.

Players that play the center position often have inflated FG%. Shaquille O’Neal, one of the best centers ever, has a career FG% of 58.1%. By comparison, Steve Nash, currently one of the top five guards in the NBA, boasts a career FG% of 48.5%. Using FG% as a benchmark, we would be led to believe that Shaquille O’Neal is a better shooter than Steve Nash, a claim that any basketball expert would regard as false. Part of the problem lies in the inability of the statistic to capture the shooting patterns for each player over different distances from the basket. We attempt to generate two unique measures that (i) remove this potential bias, (ii) capture a player’s total offensive ability, and (iii) still have an intuitive interpretation.

Researchers have recently begun to create better measures of offensive ability. One of the more notable statistic is called Offensive Rating, or ORtg. ORtg, which was invented by Dean Oliver (3), captures the number of points scored by a player per 100 possessions. Another commonly employed rating is known as Player Efficiency Rating, which was developed by John Hollinger (4). Hollinger claims thats “PER sums up all a player’s positive accomplishments,

\(^1\)In this paper, we define a field goal as a basket scored on any shot or tap other than a free throw, layup, or dunk.

\(^2\)We verified this assumption empirically by observing the shot success percentage curves for the entire NBA.
substracts the negative accomplishments, and returns a per-minute rating of a player’s performance.” While these approaches produce better estimates of a player’s ability to shoot and score than FG%, there are some major drawbacks. These statistics incorporate very little to no relativity; that is, it is difficult to make player comparisons on a positional or shooter-type level. Also, there exists a nontrivial amount of measurement error in both. These statistics are based on estimates of missing events such as number of possessions. Finally, and most importantly, the same bias that exists in FG% is a factor in the determination of ORtg.

Previous studies have done similar work that quantify shooting and offensive ability using spatial data. The first such study analyzed the performance of Michael Jordan (1). Their model considered each shot chart as an instance of some Poisson process and estimated the corresponding nonparametric functions relating to each event. The other study focused on Sam Cassell (2). Sam Cassell is known for his shooting preference, which is from the left side of the basket. A Bayesian multivariate logit model was applied to spatial data along with an added set of covariates. To determine the model parameters, sampling is done via a Monte Carlo Markov Chain method. The results from these studies proved to be helpful examples of the capabilities of this type of approach, as well as instructive for teams defending either player. While we are not addressing matters related to the defensive side of basketball in this paper, this is precisely the type of analysis we are looking to replicate. The underlying difference is that we want to make inference on the shooting ability of every NBA player in a parsimonious way.

This paper is organized into five sections. The next section is a description of the data set we used with explanations of the data’s origin and all of its attributes. Section III defines the methodology we applied to produce our measures. In section IV, we present the results, including lists of players who do perform well or poorly offensively according to our measures. The last section is a discussion of our conclusions.

2 Data

The aforementioned research on this subject (1)(2) had two dimensional data available to them. These projects used images from game recaps on sports websites called “shot charts.” These charts display a player’s or team’s shot attempts and successes mapped on top of an image of a basketball court.

Ideally, for our study, we would like to observe these shot charts for all players in as many games as possible. However, shot charts are available only
as Flash objects embedded in web pages, so the method for gathering this data is copying each shot chart down by hand for that player in a season (1), (2). Over the course of the three seasons between 2004 and 2007, more than 600 NBA players appeared in at least one game. Because we do not have the man hours available to record that much data, we use the “Play-by-Play” section for every game on ESPN.com.

We extract from ESPN.com every NBA game that occurred during the 2004-2005, 2005-2006, and 2006-2007 seasons. The resulting sample we obtain consists of 724,199 events. Theses events include jump shots (or field goals), layups, dunks, fouls, etc. and some defensive events such as steals and rebounds. In this paper, our focus is on all of the offensive events that took place during a game.

Not all of these instances are properly observed. Some of them are counted as missing because there is no distance recorded for that event. The proportion of our sample that is missing is about 21.4%. While this number may seem high, our sample is still left with 568,951 events, which is plenty of data for our analysis. The remaining concern is that a systematic pattern to the missingness might be present. We test its validity by randomly sampling 100 games. Each missing event from the 100 games is matched to that same event on that game’s shot chart. After analyzing the comparable event’s on the shot charts, we did not find any evidence that the two populations, missing and non-missing, are different, so we are confident that no biases are introduced in our calculations by dropping the missing observations from the dataset.

3 Methodology and Measures

The shot-by-shot level resolution of our dataset allows us to control for the glaring confounder in the traditional FG% statistic, which is distance. The different types of shots we consider for our analysis are field goals, dunks, layups and free throws. Field goals are further separated by distance while dunks, layups and free throws do not have a distance associated with them. By breaking down the shooting characteristics of players by distance and aggregating over 3 entire seasons, we come up with two novel and distinct measures for offensive ability: SCAB and SHTAB, which are short for Scoring Ability Above Baseline and Shooting Ability Above Baseline, respectively. The two

---

3A few games in each of the seasons were missing the “Play-by-Play” page altogether. However, there are not enough matches missing to make this a problem.

4The exact proportions of missing data for the 2004-2005, 2005-2006 and 2006-2007 seasons are as follows: 36.5%, 20.5% and 7.6%.
measures are defined as follows.

- The **SCAB** value for a player is the number of points scored by that player net the number of points scored by their comparable *baseline player* per 36 minutes of playing time with both attempting the same total number of shots according to their respective shot attempt curves of field goals, layups, dunks, and free throws.

- The **SHTAB** value for a player is a measure of his field goal shooting efficiency accounting for distance net the field goal shooting efficiency for their comparable *baseline player* accounting for distance, where distance is accounted for and both “players” have the same shot attempt curve.

The *baseline player* mentioned in the two definitions above sets a common comparison measure for players corresponding to their shot attempt profile. We discuss the concept of the baseline player and address two different ways of defining a baseline player in section 3.1.

SCAB and SHTAB measure different aspects of a player’s offensive ability. Intuitively speaking, the SCAB value of a player is a measure of the number of points a player will score, on average, above/below baseline performance per game. His SHTAB value is a measure of the number of points he will score, on average, per field goal attempt above/below baseline performance. When comparing two players A and B, if A has the better SCAB value then he is better than B on a scoring-per-minute basis, after controlling for the total number of shots taken. If A has the better SHTAB value then he is better than B on a shot-by-shot basis. A manager would prefer a player with a high SCAB to get more playing time, while a player with a high SHTAB would be preferred to come in and shoot in a situation with little time remaining or when a specific shot is crucial to a possession.

A higher SCAB value does not imply a higher SHTAB value and vice-versa, although they are correlated\(^5\). As a toy example, let’s neglect baseline performance (i.e. a baseline of 0) and consider all shots to be two-point field goals from 15 feet. Say player A attempts 20 shots in 10 minutes of playing and makes only 5 of them, which translates into 10 points. His SCAB is 36 and his SHTAB is 1.0. Now, say player B attempts 15 shots in 30 minutes of playing and makes 10 of them resulting in 20 points. His SCAB is 24 and his SHTAB is 1.3. Player A has a much higher SCAB, but player B has a much higher SHTAB.

---

\(^5\)An appropriate analogy is the relationship between batting average and on-base-percentage in baseball.
3.1 Establishing the Baseline Player

The baseline player is a hypothetical player for each player type who, for every field goal attempt over the range of distances and for every other shot type, shoots at their respective league average success percentages. The player type for each player can be defined in a multitude of ways and we detail two such ways in this section.

The concept of a baseline player is natural, since a general manager is often making a choice between two or more players who play the same position or style; a comparison of their SCAB and SHTAB values will nullify the effect of the baseline. Let's say, for example, that the general manager has to choose between players A and B playing guard. If player A has a SHTAB of +1.25 and player B has a SHTAB of +0.85, then, in terms of shooting ability, player A is the better choice because both of them have been compared to the same baseline player. This idea can also be extended in situations where a trade needs to be evaluated. Essentially, the general manager can employ these measures to give them an idea on the gains and losses of playing some player over giving the job to the league average player.

We look into two different methods for defining player type. The first method is to define it by position. A player can be one of the following: guard, guard-forward, forward, center-forward or center. The performance of the baseline player at each of these five positions is the league average performance over all players at the corresponding position.

The second method is to define them by shot attempt profile. We say that two players are of the same type if they attempt a similar number of field goals at the various distances. We use this as our definition of comparison since we are primarily interested in a player’s shooting ability. We are not interested in other methods of classification such as discerning players by how they play defensively, who they get matched up with man-to-man, or whether they call the plays. Thus, we are going to assume that these characteristics have no causal link to the success and distribution of shooting across distances.

Practically, this kind of labelling can be achieved by running a clustering procedure over the shot attempt curves, or shot distributions, of players. For a given player, their baseline comparison will be the average performance over all players in his cluster. Even though we expect most players to cluster according to their positions, we still find notable exceptions where players that were labelled as one position play in the style typical of another.
3.2 Clustering Players

We use the standard clustering procedure k-means to categorize players by their shot distributions. We initialize the method to find five centers. The intuition behind this choice relates to the number of players on court and the number of possible positions, which is five. Since the concern is mislabelling, we conjecture that a mislabelled player would be one of the other four possible types of players.

![Cluster 1](image1)
![Cluster 2](image2)
![Cluster 3](image3)
![Cluster 4](image4)
![Cluster 5](image5)

**Figure 1:** The centers of the five clusters of players found via the k-means algorithm according to each player's shot distribution.

The input for the k-means algorithm (5) is a standardized version of each player’s shot distribution. The resulting centers are encouraging, reinforcing our intuition of setting up five clusters. As seen in figure 1, each cluster represents a group of players primarily shooting at five distinct distances away from the basket. To link this back to the traditional basketball positions, consider the cluster associated with the center whose mode is at the smallest distance to the basket. It contains players who are generally thought of as centers, such as Alonzo Mourning and Shaquille O’Neal. It is easy to believe that these players shoot very similarly because they are generally positioned around the basket. An example of a player who may have been misclassified by the classical groupings is Rajon Rondo. He is often thought of as a guard,
but, according to our clustering, his shooting style is most similar to players in cluster 4, whose mode is the second closest to the basket. The players listed in cluster 4 are traditionally thought of as center-forwards.

**Table 1:** Position vs. Cluster Placement of Players

<table>
<thead>
<tr>
<th></th>
<th>G</th>
<th>G-F</th>
<th>F</th>
<th>F-C</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>82</td>
<td>17</td>
<td>42</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>89</td>
<td>22</td>
<td>52</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Cluster 3</td>
<td>20</td>
<td>7</td>
<td>35</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>Cluster 4</td>
<td>4</td>
<td>3</td>
<td>32</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>Cluster 5</td>
<td>0</td>
<td>0</td>
<td>18</td>
<td>5</td>
<td>19</td>
</tr>
</tbody>
</table>

To illustrate that there is in fact a difference between our clustering and the traditional position labelling, we display the distribution according to position for each cluster in table 1. It should be noted that, while the distributions across positions for cluster 1 and cluster 2, or similarly for cluster 3 and cluster 4, are close to the same, it is important that we do not shrink the number of clusters. Evidence against this idea can be seen in figure 1. From this figure, we can see that, for example, cluster 1 players are choosing to most often take three point shots. The mode of the shot distribution for cluster 2 is a few steps in front of the three point line. This distinction is vital to our calculations, specifically SCAB. In the case with SCAB, the baseline comparison is made not only to the average shooting curve, but also according to the cluster’s shot distribution. Because there are significant differences from cluster to cluster, it is important to use five centers and not reduce this any further.

### 3.3 Calculating SCAB

We begin by considering only field goals and add in free throws, dunks and layups in the second step. Let

- $\Delta_d = 1, 2, ..., D$: set of distances over which players attempt shots.
- $n_{f_{i,j}}$: number of field goals attempted by player $i$ and distance $j$.
- $total_{f_i}$: total number of field goals attempted by player $i$.
- $s_{i,j}$: field goal, shooting success percentage (call it success percentage) of player $i$ at distance $j$. 

Published by Berkeley Electronic Press, 2010
• $s_j$: success percentage of baseline player at distance $j$.
• $\overline{\text{prob}}_j$: probability of the baseline player shooting at distance $j$.
• $p_j$: value of a field goal shot from distance $j$.

The field goal SCAB of player $i$ is defined as

$$SCAB_{FG}(i) = \sum_{j=1}^{D} p_j(n_{f_i,j} \cdot s_{i,j} - totalf_i \cdot \overline{\text{prob}}_j \cdot s_j)$$

$SCAB_{FG}$ is a measure of a player’s scoring ability compared to a baseline player, who takes the same total number of shots, but according to the baseline shot distribution, not player $i$’s shot distribution. We choose this difference in shot distributions because we expect that a player at a position or in a cluster will tend towards shooting like the average player for that position/cluster.

Dunks, layups, and free throws do not have distances associated with them. We define the Dunks SCAB of player $i$ as

$$SCAB_D(i) = (d_i - \overline{d}) \cdot nd_i \cdot 2$$

where $d_i$ is the success percentage of dunks by player $i$, $\overline{d}$ is the success percentage of dunks by his/her baseline player and $nd_i$ is the number of dunks attempted by player $i$. The factor of 2 is included because each dunk is worth 2 points. We similarly define $SCAB_L$ for layups and $SCAB_{FT}$ for free throws for each player. Note that the constant of 2 remains for layups but disappears for free throws as each free throw is worth only 1 point. Combining the above defined measures for each shot type, we get the SCAB value for player $i$ as

$$SCAB(i) = (SCAB_{FG}(i) + SCAB_D(i) + SCAB_L(i) + SCAB_{FT}(i)) \cdot \frac{36}{Mins_i}$$

where $Mins_i$ is the number of minutes played by player $i$. The factor of 36 appears because it is the average number of minutes played by a starter in the NBA.

### 3.4 Calculating SHTAB

The Points Value (PV) of player $i$ at distance $j$ is calculated as

$$PV(i, j) = (s_{i,j} - \overline{s}_j) \cdot n_{f_i,j} \cdot p_j$$
Intuitively, $PV(i, j)$ tells you the total number of extra points player $i$ scores when compared to their baseline player, if that baseline player had taken the exact same field goal attempts that player $i$ took. The SHTAB of player $i$ is calculated as his Points Value for field goals aggregated over all distances divided by their number of attempted field goals.

$$SHTAB(i) = \frac{\sum_{j=1}^{D} PV(i, j)}{\sum_{j=1}^{D} n f_{i,j}}$$

We exclude dunks, layups and free throws from this statistic because these shots do not represent the “shooting ability” of a player and the contribution of these shots are implicitly accounted for in SCAB. For example, if a player is bad at free-throws then the defense can consistently foul him and get him to the free throw line. As a result, his/her SCAB value will suffer.

### 3.5 Smoothing Field Goal Success Percentage Curves

We now address two issues with the success percentage curves for each player (i.e. the curves defined by $s_{i,j}$ for $j \in \Delta_D$). These curves give the empirical success percentages a given player has achieved with field goal attempts over various distances. The first issue is that the field goal shooting ability of player $i$ at distance $j$ is clearly not independent of his ability at distances $j - 1$ and $j + 1$. More generally, we need to acknowledge the fact that there is dependence between a player’s shooting ability over various distances and that this dependence fades with increasing gaps in distances.

The second issue derives from the number of shot attempts made by the player at distance $j$. If the number of attempts is too few, then the resulting empirical success percentage would be extremely noisy. In the extreme case, say that a player attempts only one shot at a given distance. His success percentage at that distance will be 1.0 if he makes the shot and 0.0 if he misses. Hence, some sort of compensation is needed to derive a more accurate estimate of his success percentages at distances with too few shot attempts.

Kernel smoothing lends itself as a natural solution to the first issue. A real-valued function $K(u)$ is a kernel if $\forall u \in \mathbb{R}$

1. $K(u) \geq 0$
2. $K(u) = K(-u)$
3. $\int_{-\infty}^{+\infty} K(u) du = 1$
The kernel smoothed estimate of success percentage of player $i$ at distance $j$ is given by

$$\hat{s}_{i,j} = \frac{\sum_{u=-\delta}^{+\delta} K(u) \ast s_{i,j+u}}{\sum_{u=-\delta}^{+\delta} K(u)}$$

where $2\delta$ is the width of window in feet over which smoothing is being performed. We use the Gaussian kernel as our kernel function for the smoothing estimator. It is defined as

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$$

To deal with the second issue, the kernel smoother described above is modified to account for the number of shot attempts at the various distances. The kernel weights are multiplied by the respective number of shot attempts at that distance and the normalization constant is recalculated. Our new modified kernel smoother is defined as the following:

$$\hat{s}_{i,j} = \frac{\sum_{u=-\delta}^{+\delta} (K(u) \ast s_{i,j+u} \ast n_{f_{i,j+u}})}{\sum_{u=-\delta}^{+\delta} (K(u) \ast n_{f_{i,j+u}})}$$

This final $\hat{s}_{i,j}$ is the estimate of player $i$’s true success percentage that we use when obtaining player $i$’s SCAB and SHTAB.

### 3.6 Refining the 2-pt/3-pt Criterion

The three point line on the NBA court forms an arch on the floor around the basket. Some points on the line are as close as 22 feet to the basket, while, at the top of the arch, the basket is 23 feet 9 inches away. Due to this variation, there is an area of ambiguity when classifying whether a shot at a given distance is a two pointer or three pointer.

Instead of imposing a cut off or outside restraint to when a basket should be considered a three pointer, a logit curve was fit to the data. The resulting curve (see figure 2) is an estimate of the proportion of three point shots made at some distance away from the basket. The logit model appears to be a great choice in this setting, since the proportion of three pointers stays at 0 until around 22 feet where it rapidly increases to 1 by around 24 feet. With these estimates, we can assign more accurate values to the $p_j$’s that are used when finding SCAB and SHTAB.
4 Results

All of the calculations are tabulated over the three seasons of data. Constraints on the number of shots a player has taken are put in place to prevent small sample size issues\textsuperscript{6}. The following subsections detail the specific results relating to our different methods of calculation.

4.1 SCAB/SHTAB by Position

One point that should be made, which will appear in all of the proceeding tables, is that there is little intersection between the two measures, which is expected. Since defense is not taken into account in our methods, we would suspect that the SHTAB leaders would be players who do not get as heavily defended but are, nonetheless, great shooters. As for SCAB, these top players

\textsuperscript{6}The constraint is 30 shots. This number is chosen because it eliminated one third of all NBA players, a convenient amount. Any more players may remove significant one season players, while any less may allow for small sample aberrations.
Table 2: The top and bottom SCAB/SHTAB players using their raw success curves and their respective position baseline.

are credited for taking more attempts. These shooters may not be the best shots, but they are able to generate many good chances. There are exceptions to these generalizations, with the most significant anomaly being Steve Nash. He defies said logic by consistently being both a great shooter and a creator of good chances.

Table 2 displays the top/bottom 10 SCAB and SHTAB players using the method that takes a player’s raw curve and compares them to the baseline player at their position. Besides a few big names in the top ranked players, most of those listed are not well-known for their shooting or offensive abilities,

\[7\]It should be noted that his ability to make chances is influenced by his supporting cast. Regardless, his prevalence at the top of the tables listed in this paper is a testament to his skills as a shooter.
so it seems natural to modify this calculation by employing a different baseline for comparison.

4.2 SCAB/SHTAB by Clustering

<table>
<thead>
<tr>
<th>Name</th>
<th>SCAB</th>
<th>Name</th>
<th>SHTAB</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Best Ten Players</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wesley Person</td>
<td>2.58</td>
<td>Fred Hoiberg</td>
<td>0.300</td>
</tr>
<tr>
<td>Dirk Nowitzki</td>
<td>1.99</td>
<td>Danny Fortson</td>
<td>0.246</td>
</tr>
<tr>
<td>Glenn Robinson</td>
<td>1.98</td>
<td>Zeljko Rebraca</td>
<td>0.246</td>
</tr>
<tr>
<td>Walter Herrmann</td>
<td>1.84</td>
<td>Steve Nash</td>
<td>0.224</td>
</tr>
<tr>
<td>Ben Gordon</td>
<td>1.78</td>
<td>Wesley Person</td>
<td>0.204</td>
</tr>
<tr>
<td>Jason Terry</td>
<td>1.77</td>
<td>Matt Bonner</td>
<td>0.184</td>
</tr>
<tr>
<td>Steve Nash</td>
<td>1.74</td>
<td>Walter Herrmann</td>
<td>0.183</td>
</tr>
<tr>
<td>Fred Hoiberg</td>
<td>1.68</td>
<td>Jason Terry</td>
<td>0.175</td>
</tr>
<tr>
<td>Eddie House</td>
<td>1.58</td>
<td>Robert Swift</td>
<td>0.170</td>
</tr>
<tr>
<td>Raja Bell</td>
<td>1.49</td>
<td>Glenn Robinson</td>
<td>0.166</td>
</tr>
<tr>
<td><strong>Worst Ten Players</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dajuan Wagner</td>
<td>-4.19</td>
<td>Michael Ruffin</td>
<td>-0.600</td>
</tr>
<tr>
<td>Earl Barron</td>
<td>-3.92</td>
<td>Renaldo Balkman</td>
<td>-0.468</td>
</tr>
<tr>
<td>Nikoloz Tskitishvili</td>
<td>-3.87</td>
<td>Tale Davis</td>
<td>-0.389</td>
</tr>
<tr>
<td>Awvee Storey</td>
<td>-3.85</td>
<td>Vassilis Spanoulis</td>
<td>-0.388</td>
</tr>
<tr>
<td>Yaroslav Korolev</td>
<td>-3.85</td>
<td>Kevin Burleson</td>
<td>-0.344</td>
</tr>
<tr>
<td>Terence Morris</td>
<td>-2.85</td>
<td>Nikoloz Tskitishvili</td>
<td>-0.342</td>
</tr>
<tr>
<td>Robert Hite</td>
<td>-2.63</td>
<td>Dee Brown</td>
<td>-0.325</td>
</tr>
<tr>
<td>Brandon Bass</td>
<td>-2.60</td>
<td>Awvee Storey</td>
<td>-0.321</td>
</tr>
<tr>
<td>Elden Campbell</td>
<td>-2.60</td>
<td>Tyrus Thomas</td>
<td>-0.314</td>
</tr>
<tr>
<td>Kevin Burleson</td>
<td>-2.42</td>
<td>Lonny Baxter</td>
<td>-0.309</td>
</tr>
</tbody>
</table>

Table 3: The top and bottom SCAB/SHTAB players using their raw success curves and their respective cluster baseline.

The differences between the values calculated using the cluster baseline (see table 3) and those found using positional baseline are fairly large. More importantly, we see significant changes to the appropriate players. For example, the 5 of the top 10 SCAB players using the positional baseline are in the top 10 SCAB players using the cluster baseline, with Dirk Nowitzki, Ben Gordon, Eddie House, and Raja Bell all seeing significant boosts, while replacing...
some relatively unknown players such as Anthony Roberson and Travis Diener. Since the resulting rankings from clustering make sense in our setting, we continue to employ them as our baseline. The final step is to smooth the individual shooting curves and shrink them to the cluster means.

### 4.3 Final SCAB and SHTAB

<table>
<thead>
<tr>
<th>Name</th>
<th>SCAB</th>
<th>Name</th>
<th>SHTAB</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Best Ten Players</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dirk Nowitzki</td>
<td>1.75</td>
<td>Steve Nash</td>
<td>0.161</td>
</tr>
<tr>
<td>Ben Gordon</td>
<td>1.42</td>
<td>Jason Terry</td>
<td>0.137</td>
</tr>
<tr>
<td>Jason Terry</td>
<td>1.40</td>
<td>Dirk Nowitzki</td>
<td>0.117</td>
</tr>
<tr>
<td>Raja Bell</td>
<td>1.28</td>
<td>Elton Brand</td>
<td>0.102</td>
</tr>
<tr>
<td>Eddie House</td>
<td>1.25</td>
<td>Ben Gordon</td>
<td>0.100</td>
</tr>
<tr>
<td>Steve Nash</td>
<td>1.17</td>
<td>Zeljko Rebraca</td>
<td>0.098</td>
</tr>
<tr>
<td>Elton Brand</td>
<td>1.03</td>
<td>Matt Bonner</td>
<td>0.095</td>
</tr>
<tr>
<td>Mike Miller</td>
<td>1.00</td>
<td>Raja Bell</td>
<td>0.094</td>
</tr>
<tr>
<td>Wesley Person</td>
<td>0.99</td>
<td>Mike Miller</td>
<td>0.092</td>
</tr>
<tr>
<td>Damon Jones</td>
<td>0.98</td>
<td>Brian Cook</td>
<td>0.084</td>
</tr>
<tr>
<td><em>Worst Ten Players</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brandon Bass</td>
<td>-1.63</td>
<td>Ben Wallace</td>
<td>-0.141</td>
</tr>
<tr>
<td>Damir Markota</td>
<td>-1.44</td>
<td>Josh Smith</td>
<td>-0.116</td>
</tr>
<tr>
<td>Yaroslav Korolez</td>
<td>-1.38</td>
<td>Michael Ruffin</td>
<td>-0.110</td>
</tr>
<tr>
<td>Charles Smith</td>
<td>-1.30</td>
<td>Desmond Mason</td>
<td>-0.100</td>
</tr>
<tr>
<td>Stephen Graham</td>
<td>-1.21</td>
<td>Dale Davis</td>
<td>-0.098</td>
</tr>
<tr>
<td>Josh Powell</td>
<td>-1.20</td>
<td>Emeka Okafor</td>
<td>-0.090</td>
</tr>
<tr>
<td>Earl Barron</td>
<td>-1.08</td>
<td>Jeff Foster</td>
<td>-0.085</td>
</tr>
<tr>
<td>Roger Mason</td>
<td>-1.06</td>
<td>Eddie Griffin</td>
<td>-0.084</td>
</tr>
<tr>
<td>Nikoloz Tsikitishvili</td>
<td>-1.04</td>
<td>Jamaal Magloire</td>
<td>-0.082</td>
</tr>
<tr>
<td>Pat Burke</td>
<td>-0.98</td>
<td>Andrei Kirilenko</td>
<td>-0.076</td>
</tr>
</tbody>
</table>

*Table 4:* The top and bottom SCAB/SHTAB players using their smoothed success curves and their respective cluster baseline.

In table 4, we have our final calculations of SCAB and SHTAB calculated by clustering, smoothing, and shrinking. The SCAB top 10 list contains all well-known players, except for Damon Jones, a veteran bench player. Even better, both top and bottom SHTAB rankings consist of players that are recognizable, whose respective order is conceivable. No longer do we see players
associated with a small sample of shots popping up into our top 10 (or bottom 10) list as before. All of the players in table 4 have seen enough playing time that they warrant their spot in the rankings. An important improvement that comes with using this final method is the resulting magnitudes of SCAB and SHTAB. They seem much more reasonable than those numbers found doing the previous calculations. For example, we would not expect that, even when comparing the best and worst players, the difference in SHTAB’s of two shooters is close to one point per field goal attempt. That is an artifact of the roughness in the empirical shooting curves, not a truth about the difference in the shooting abilities of the best and worst players in the NBA.

Figure 3 shows a scatter plot between the final SCAB and SHTAB values for all players broken down by cluster. The overall plot contains the 95% and 99% data ellipses, which serve to approximate the joint distribution of these statistics. Note that points to the top right of the plot are clearly the best players and points to the bottom left are the worst. Cluster 3 shows that Ben Gordon and Jason Terry are significantly better than the others in their cluster (i.e. their points lie outside the 99% data ellipse), while for clusters 4 and 5, the best players are Steve Nash and Dirk Nowitzki, respectively.

We attach names to some of the interesting points on the previous scatter plot, as well as label points corresponding to some popular players in figure 4. The best players in the league, according to our measures, turn out to be Steve Nash, Dirk Nowitzki, and Jason Terry. Shaquille O’Neal appears to be only an average shooter and scorer over this three year span using our statistics, just as we contended in the introduction of this paper. The location of LeBron James might seem surprising to some people. However, considering that he often gets double-teamed, is forced to shoot from unfavorable positions, and has only average talent surrounding him, his location on this plot can be justified. Quincy Douby is an average field goal shooter, but scores high on SCAB due to his extremely high free throw percentage.

### 4.4 Comparison to Existing Measures

Contrasting our measures to existing measures of offense in basketball is an important method of external validation. Figure 5 illustrates these juxtapositions. In it, we plot SCAB against ORtg, OWS and PER, and SHTAB against EFG%, TS% and 3P%.\(^8\) We also include table 5 of simple correlations describing in a numerical sense the relationship between our measures and the extant ones.

\(^8\)For an explanation of each statistic, see Appendix
Figure 3: Scatter plot of SHTAB vs. SCAB values for all players together and broken down by cluster.
Figure 4: Locations of a few big name players on the scatter plot of SHTAB vs. SCAB along with a few other interesting points.
Figure 5: Plots comparing SCAB and SHTAB to their respective counterparts. For SCAB, those include ORtg, OWS, and PER, while SHTAB is compared to its contemporaries EFG%, TSP, and 3P%.
Table 5: Two correlation tables for the various offensive metrics. These measures are split into two groups, one for overall measures of offense (i.e. SCAB) and one for field goal specific measures (i.e. SHTAB).

We begin by noting that SCAB is reasonably (and positively) correlated with all of the relevant, advanced statistics of offense. There is an exception with OWS; a correlation is present, but by no means a strong one. Still, all of these correlations are weak enough to indicate that there is a distinct difference amongst them. We expand more on this distinctness below.

The definition of Offensive Rating goes along the same lines of SCAB. Because it is unclear whether the number of possessions is a good, base metric, though, SCAB may be a much more accurate measure of an individual’s ability to score points. There might be multiple possessions in which the player might not be involved at all and many in which the player’s involvement is minimal, with the ball simply having passed through his/her hands. PER or Player Efficiency Rating is a measure of the overall value of that player, which includes offense, defense and turnovers. The positive correlation between PER and SCAB would indicate that a player who performs well at offense can be expected to be better in terms of overall efficiency. Unlike our methods, the calculation of PER is rather complicated and involved, assigning subjective weights to different player related events in a game. SCAB takes a purely objective view of scoring points as a function of time, as long as employing a proper baseline. Finally, it is expected that Offensive Win Shares is going to deviate and differ from SCAB. OWS is more a measure of how a player contributed to a win, not how much a player contributed overall offensively.

SHTAB is compared to Effective Field-Goal Percentage, True Shooting Percentage and 3-Point Percentage, all three of which are measures of shooting efficiency on a per-shot basis. We found that SHTAB is uncorrelated with all of...
Table 6: The SCAB and SHTAB, as well as more traditional, advanced statistics, of the pre- and post-trade seasons for the two players discussed below.

these measures, which draws on the reality that none of these existing measures account for distance of shooting. The introduction of distance into shooting ability causes players such as Shaquille O’Neal, who score well on EFG% to drop down into the middle of the pack in SHTAB. In fact, it is encouraging to note that by simply being good three point shooter does not necessitate great shooting ability, as seen in its near-zero correlation to SHTAB.

4.5 Case Studies

We provide two case studies of players who have recently been traded for the purpose of highlighting our new measures’ stability and predicability. Their pre- and post-trade statistics (i.e. year before being traded and year after being traded) are displayed in table 6. The first such player is J.R. Smith, who is currently considered to be a good young player at the guard position. His career did not begin all that inspired, however. After a below-average freshman campaign, he did nothing to inspire future greatness the following season. According to his traditional metrics, he failed to significantly improve on his 2004-2005 numbers, especially with regard to PT/MP and EFG%. His employer at the time, the New Orleans/Oklahoma City Hornets, traded him away as a result. The Nuggets took this opportunity to upgrade at the guard position, knowing that there was hidden potential in Smith’s game. In his first season with Denver, Smith saw all of his statistics increase, with dramatic progressions in PT/MP, EFG%, and TS%. By looking at his 2005-2006 SCAB and SHTAB calculations, it was clear that he was at worst an average player at his position, a fact that was not captured by other metrics. Moving onto the 2006-2007 season, we observe nearly no change in the numbers, but we find his other statistics matching up with his SCAB/SHTAB evaluation. This
proved to be quite the haul for the Nuggets, who gave up little in terms of value to acquire Smith’s services.

Bobby Simmons’ recent move from Milwaukee to New Jersey is another case where SCAB and SHTAB predicted stable performance across seasons while classical measures proceeded to undervalue his efforts in his pre-trade season. While his SHTAB and SCAB values remained nearly the same across both seasons, his EFG%, TS% and ORtg numbers improved significantly after the trade. Both his SCAB and SHTAB values already showed him to be an above-average league player; this fact that was not supported by his pre-trade season, but his post-trade performance did support it. Once again, knowing that Simmons was undervalued by his own team, the Nets were able to not only acquire him, but also the promising player Yi Jianlian in an affordable package.

This is only two of the many examples where player’s SCAB and SHTAB calculations remain constant across seasons and their traditional statistics fluctuate. Cases in which player’s SCAB and SHTAB values vary do exist, but to a much lesser extent than their traditional contemporaries. For the purposes of brevity, we limit ourselves to the two examples mentioned above, as evidence of the relative stability and predicatability of SCAB and SHTAB across seasons.

5 Conclusion

In this paper, we have outlined two new measures of offensive ability in basketball, with both statistics accounting for the distance of the shot from the basket: SCAB, which details how much more/less a player scores above/below the baseline as a function of time on court, and SHTAB, which describes how good/bad the player is at shooting field goals. Both measures have intuitive interpretations and can be used to effectively rank basketball players by shooting ability and quantify the offensive differences between them. A manager looking to optimally allocate his players can use SCAB to determine the length of time that each player spends on court and SHTAB to decide who amongst those on the court should be given the ball to shoot in certain situations. Most importantly, a general manager will be able to evaluate shooters based on more unbiased measures, allowing them to better fill in roster weaknesses and buy low on undervalued players.

While the use of SCAB and SHTAB lead to legitimate rankings of players, there are a few issues that need to be considered while doing evaluations based on them. They can be affected by the way opposition defend against
the player in question and by the support on offense that the player gets from his teammates; these issues may factor into a player’s final SCAB and SHTAB ratings. Some players may be perform well at creating good shooting opportunities for themselves while other players may be getting easier shooting opportunities. This could be a direct result of the defense concentrating on a more formidable teammate or the vision and play-creating skills of an assisting teammate. Likewise, a big name player such as LeBron James may find himself being double-teamed often and as a result, may end up having to shoot from unfavorable positions. These factors vary on a game-to-game and, possibly, a quarter-to-quarter basis making it difficult to adjust accordingly. On a more superficial level, it is common knowledge that teams defend more against a player taking a 20-foot shot at the top of the key than they would against that same player taking the same shot along the baseline. Both of these shots are taken at the same distance, but one will be more difficult when regarding angle to the basket. We believe that aggregating over all three seasons will lead to an averaging out of such effects.

Future work along these lines will move towards analysis of defenses and interactions of defenses with our two measures of offense. Our dataset will allow us to characterize the strength of a defense in terms of the distribution of shots made against it. By interacting player’s shot distributions with opposition’s defensive shot distributions, one could potentially find out how a player performs against good or bad defenses. Often players are labelled as flat-track bullies or as clutch performers and this methodology could provide a way to quantify such tags.

Appendix: Other Offensive Measures

- 3P%: Three point field goal percentage.

- EFG%: Effective Field Goal percentage; this metric controls for the differing value between a three point and two point field goal. The equation to calculate the statistic is \( \frac{FG + 0.5 \times 3P}{FGA} \).

- ORtg: Offensive Rating; this represents the number of points produced per 100 possessions. It was originally invented by Dean Oliver (3).

- OWS: Offensive Win Shares; similarly calculated to Win Shares in baseball, which were invented by Bill James (6), except for basketball. Essentially, it measures the number of wins a player contributes to offensively when compared to the league.
• **PER**: Player Efficiency Rating; it is a rating created by John Hollinger (4). According to Hollinger, it incorporates “all [of] a player’s positive accomplishments, subtracts the negative accomplishments, and returns a per-minute rating” in terms of a player’s performance.

• **TS%**: True Shooting Percentage; it measures the shooting efficiency of a player by accounting for two point and three point field goals, as well as free throws. Its equation is $\frac{PTS}{2 \times (FGA + 0.44 \times FTA)}$.

**References**


