Distribution Agreement

Kyle Monk **April 4, 2017**

How Should High Volume Scorers in the NBA be Valued?

by

Kyle Monk

Christina DePasquale

Advisor

Department of Economics

Christina DePasquale

Advisor

Yunxiao Chen

Committee Member

Arri Eisen

Committee Member

2017

How Should High Volume Scorers in the NBA be Valued?

by

Kyle Monk

Christina DePasquale

Advisor

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Department of Economics

2017

Abstract

How Should High Volume Scorers in the NBA be Valued?

By Kyle Monk

Many basketball metrics have started to focus on efficiency. Unfortunately, the emphasis on efficiency comes with a tradeoff. Just as there is a constant bias-variance tradeoff in econometrics and modeling, there is an efficiency-volume tradeoff in basketball. I will propose the creation of a statistic called Cumulative Marginal Possessions (CMP) to help evaluate the efficiency-volume tradeoff. This metric is built from the assumption that every point scored in the NBA can be accounted for by the offense (plays, positioning, timing, etc.), defense (schemes, positioning, aggressiveness, etc.), and the ability of the players involved. The metric will evaluate player ability by modeling the outcome of each possession using a non-parametric model of the form:

$$Points = B_0 + B_1 Potential\ Assist + f_1(shot\ clock) + f_2(point\ differential) + f_3(time\ remaining) + \varepsilon$$

Using the actual results and predictions, the residuals of the model can assign value to the players involved in the outcome of each individual possession (efficiency). Then, these values (the player values of all possessions for a season) can be summed together to assess the entire volume of work of a player. Drawing on the foundations of the WS model (comparing points created to a league average measure), but utilizing possession data to make this distinction between the values of each possession is where CMP can begin to solve the efficiency-volume tradeoff problem. This draws on the assumption (and longtime argument of basketball traditionalists) that high volume scorers are asked to take more difficult shots- either at the end of possessions or games. Comparing the results of this analysis (using data from the 2015-2016 season) to existing metrics, like PER and WS, allows us to judge the value of CMP. Finally, comparing CMP to new contracts for free agents from the 2016 off season allows us to evaluate possible market inefficiencies in the market for free agents.

How Should High Volume Scorers in the NBA be Valued?

by

Kyle Monk

Christina DePasquale

Advisor

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Department of Economics

2017

Acknowledgements

I would like to thank Andrew Francis-Tan and Christina DePasquale for their support and development of my research idea. I would also like to thank Yunxiao Chen and Arri Eisen for their input and support. Finally, I thank my family and friends.

Table of Contents

I. Introduction

In a perfectly competitive world where each franchises' main goal is to win, a NBA franchise will pay its players equal to the value they contribute on the court. This is the ideal world popularized by the "moneyball" revolution in Major League Baseball. *Moneyball* specifically refers to the title of Michael Lewis' bestselling book about Billy Beane's attempt to win as the General Manager of the Oakland A's. Beane utilized statistics and analytics first developed by the statistician Bill James to find players who were undervalued by the league. At first, Beane found an inefficient market created by executives undervaluing on-base percentage. Taking advantage of this market inefficiency allowed him to maximize wins despite a small payroll. As the "moneyball" revolution developed, a statistic known as WAR (wins above replacement) began to dominate the landscape. This metric is designed to estimate the number of wins a baseball player will contribute to his team above what a replacement level player (defined as the highest level of minor league players) will contribute because even a team of all replacement level players will win games in baseball.

Basketball began its statistical movement by adopting many of the ideas from baseball. For example, teams have realized the greater value of three point shots and limited value in mid-range shots in an attempt to improve efficiency and points scored. However, a major difference between basketball and baseball is the distribution of outcomes. Whereas in baseball there is a defined batting order, in basketball the five players sharing the court also collectively share the shot attempts. Thus, there is a more variable distribution of shots than there are of at bats. As a result, acquiring a player in the NBA is not as simple as understanding a player is worth 5 wins above a replacement player because that player was only worth those five wins when he used his specific percentage of possessions and minutes.

To deal with this dilemma, many basketball metrics focus on efficiency. For instance, PER, developed by the Memphis Grizzlies general manager when he was working for ESPN, is a very popular statistic which is designed to summarize the positive and negative contributions that a player makes on a per minute basis ("Calculating PER," n.d.). Additionally, just as baseball began to emphasize on base percentage and slugging percentage over batting average, basketball began to emphasize metrics like true shooting percentage and effective field goal percentage over field goal percentage (the former weigh the value of the shot, 1, 2, or 3, while the latter is just the percent of shots made).

Unfortunately, the emphasis on efficiency comes with a tradeoff. Just as there is a constant bias-variance tradeoff in econometrics and modeling, there is an efficiency-volume tradeoff in basketball. For example, there is an established idea in NBA circles of an "efficiency curve" which essentially says that the more a player is asked to shoot the worse his efficiency ratings get because he can no longer be as selective. In fact, when presenting efficiency metrics it has become common practice to present individual results in two groups- all players and a more selective group of "high volume" shooters. The "high volume" category is usually defined with a given number of shot attempts per game or possessions used per game. This ad hoc distinction between player types makes it clear that there is a fundamental inconsistency in these statistics because the metrics fail to account for any volume. Unfortunately, there are few metrics that try to quantify the efficiency-volume tradeoff.

One utilized metric in the NBA that attempts to measure a season long contribution (volume) is Win Shares. This is basketball's version of WAR. Win Shares attempts to quantify the number of wins a player adds to his team. Just as WAR has three components (offense, defense, and base running), WS is made up of an offensive and a defensive component.

Offensive Win Shares are calculated based on an estimation of the points a player creates and an estimation of the number of possessions a player takes part in. Using a player's season stats, the number of points created is a weighted sum of the number points scored and the number of assists a player has. Then, the number of possessions is calculated using a weighted sum of the number of shots, free throws, assists, and turnovers committed by a player (Oliver 346-349, 2004). Using the number of points and the number of possessions, a points per possession calculation can be made. After subtracting a proportion of the league average possession (efficiency) and multiplying by possessions, the number of points added by an individual player is calculable. For the purpose of comparison, these added points are then converted to a number of wins added ("NBA Win Shares," n.d.).

There are a few things critics dislike about WS. Unlike WAR, WS is modeled from zero, not a replacement level player. In other words, a team that has WS summed to 0 would be expected to win 0 games, while a baseball team with a cumulative WAR of zero would be expected to win 48 games ("War Explained," n.d.). More importantly for the purposes of this paper, the WS metric assumes all shots occur under equivalent circumstances. In other words, the time left in the game, the score, or even the time on the shot clock are independent of the value given to a shot or possession under the WS model. However, this is an assumption that may break down for the shot selection of high volume scorers.

This paper will attempt to add insight into the value of high volume scorers by creating a metric directly targeted at the efficiency-volume tradeoff and evaluating salaries with this metric in mind. Additionally, this paper will look to analyze possible inefficiencies in the NBA labor market that have resulted from the emphasis of efficiency without equally quantifying volume.

I will propose the creation of a statistic called Cumulative Marginal Possessions (CMP). This metric will evaluate the outcome of each individual possession that a player is involved in (efficiency) and sum these values together to assess the entire volume of work of a player. Drawing on the foundations of the WS model, but utilizing possession data to make this distinction between the values of each possession is where CMP can begin to solve the efficiency-volume tradeoff problem. This draws on the assumption (and longtime argument of basketball traditionalists) that high volume scorers are asked to take the more difficult shots- either at the end of possessions or games.

The paper will be formatted in the following way. Section II will look at previous literature that can help quantify the elements of WS on a marginal value by possession basis. Sections III and IV will describe the data and methods this paper will use to actually carry out the per possession analysis. Finally, Sections V and VI will describe the results of the analysis and its possible implications for a competitive basketball market.

II. Literature Review

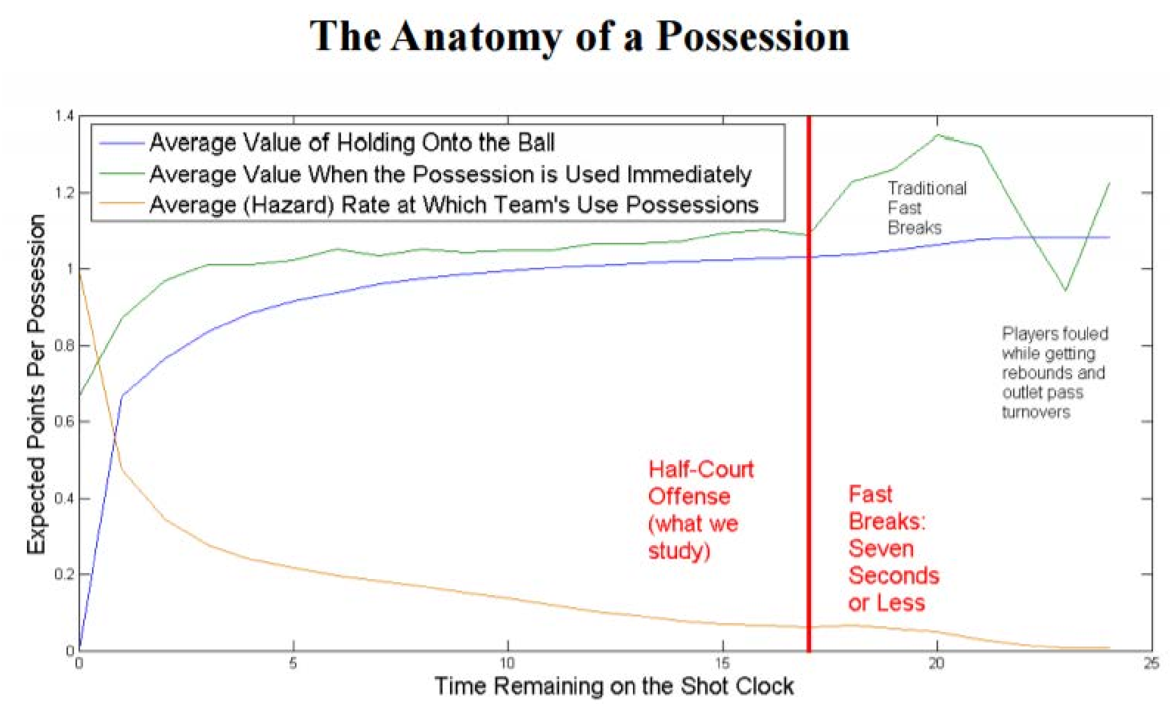
i. Salaries framed by efficiency-volume tradeoff. In the new analytics wave of reporting surrounding basketball, some experts argue that scorers can be overpaid and overrated. In the current NBA, Rudy Gay is often a player that critics will say is overrated and overpaid because he is paid mostly based on his reputation for scoring. They say he scores a lot simply because he shoots a lot and his poor efficiency is not helping his team win games. Historically, Allen Iverson (and his 26.7 ppg that required 21.8 attempts per game and came on a 42.5 field goal percentage) has faced similar criticisms. Recent literature has partially taken the side of these critics. A 2015 paper published in *The Sport Journal* studied the "Determinants of NBA Player

Salaries." The authors expected to find new advanced metrics (like PER or WS) to be deterministic of salaries. Surprisingly, they found that points scored and field goal percentage were the most statistically significant variables for determining salary (Lyons, 2015). The significances of field goal percentage and points are a perfect illustration of the efficiency-volume tradeoff. On the one hand, teams are valuing efficiency; on the other hand, simply scoring more is highly predictive of a greater salary. While both are important (especially when looked at together) some will still point to the significance of a counting stat like points as an example of the overvaluing of points. Points themselves actually do little to tell the story of the value of a player- points ignore defense, playmaking, and most importantly shot selection.

ii. Modeling possessions. In 2011, Matt Goldman and Justin Rao published and presented a paper at the Sloan Sports Conference aimed at determining how NBA players fair in decision making. They found that players, rather amazingly, almost always make the proper decision about whether to shoot or pass during a possession. NBA players have an uncanny ability to make correct decisions and shoot when the expected value of ending the possession is greater than the expected value of continuing the possession (Goldman, 2011). This is important because it implies that when looking at possessions by players we can assume that their decisions to shoot or pass are correct. Were this not the case, then high volume scorers who are constantly taking tougher shots could arguably be making poor decisions. Furthermore, it establishes that a player's decision to shoot is actually representative of the conditions and expectations of the possession.

More important than the actual results are the methods used by Goldman and Rao to analyze the decision making process. They built a model to estimate the value of continuing a possession at every moment of the possession. If the expected value of shooting was greater than the expected value of continuing the possession, then the proper decision is to shoot. This model and the cutoff values for shooting are shown in Figure 1. As you can see from looking at the green and blue lines, as the shot clock approaches 0 the value of a possession sharply decreases. The idea to model the expected value of a possession based on the shot clock opens the door for CMP to distinguish between possessions and evaluate the shots high volume scorers take relative to everyone else.

Figure 1.



Source: Goldman, 2011

In 2015, Benjamin Morris of FiveThirtyEight published an article in which he created a metric to evaluate shooting. His analysis built on Goldman and Rao's in the sense that it utilized the concept of modeling a shot expectation. Morris built a model to predict the probability of any given shot being made given the conditions of that shot. His model was a logistic regression consisting of the time remaining on the shot clock, shot distance, and distance from the nearest defender. Given the actual result of each shot taken, Morris used the residual of the model to represent the "value added" by shooting in an effort to determine the most valuable shooter (Morris, 2015). This analysis only looks at shot attempts and is a way to assess ability more than it assesses value because shot difficulty is the determinant of the expectation. However, this analysis establishes the idea of using the residuals of a model to compare players to the average.

iii. Passing as function of shooting. In addition to shooting, the Win Shares model relies on assists to calculate the number of points created. Statisticians at 82games.com attempted to add insight into value created by assists. 82games.com was a website devoted to pushing the limits of NBA statistics before tracking data became available. As a result, they hired people to chart all of the NBA games specifically looking at a statistic called a "potential assist." A potential assist is defined as any pass that leads directly to a shot, foul, or turnover (essentially any pass that occurs immediately before a possession ends). Tracking these potential assists allowed them to determine the change in result given a potential assist ("Game Charting Insights: The Value of a Good Pass," 2006). Their results are summarized in Figure 2. These results show that simply receiving a pass increases your chance of making a shot by about eight percentage points, on average. This provides a theoretical basis for CMP to include a potential assist variable in its model. Additionally, the value added from this variable should be gained by the passer rather than the shooter.

Figure 2: Measuring FG% from a "potential assist" pass versus being unassisted

Shot Type	Potential Assist	Not Assisted	Difference in FG%	Pot.Ast% of Att
3-Point	0.379	0.342	3.70%	81%
Long 2-Point	0.458	0.363	9.50%	52%
Close 2-Point	0.613	0.487	12.60%	43%
Dunks	0.91	0.84	7.00%	76%
All Shots (excluding Tips)	0.502	0.421	**8.10%**	56%

Source: "Game Charting Insights: The Value of a Good Pass," 2006 *iv. Rebounding and its value.* The final element that is incorporated into the offensive portion of the WS model is offensive rebounding (defensive rebounding is included in defensive WS). A paper presented at the 2012 Sloan Sports Conference by researchers from University of Southern California sought to determine what goes into a rebound. They published "Deconstructing the Rebound with Optical Tracking Data" using the SportVU player tracking data that records the ball and player movement at every moment of the game. An interesting finding of the research was that based solely on the location of the shot and the players at the time of the release, the outcome of the rebound could be predicted correctly 77% of the time (Maheswaran, 2012). This finding implies that the majority of an outcome of a rebound is simply the positioning established by offensive and defensive systems.

When assigning value for contributing rebounds it is important to keep in mind the idea of diminishing value of rebounds. In other words, there are some people who hypothesize that because there are only a finite number of rebounds available in a game one player getting extra rebounds could be as simple as him getting rebounds that another player on his team is already likely to get. The previous paper looking at rebounds begins to answer this question because of the fact that rebounds can be predicted very well based solely on the locations of the players. This suggests that a player gaining extra rebounds may be getting rebounds his positioning did not already predict him to get. To further study this point a contributor for countthebasket.com (an advanced basketball statistics blog) looked at several regressions comparing team rebound rates and individual player rebounding rates. The regression that included position found that across positions there are constant returns for rebounding. On average, a one percent point increase in rebound rate among point guards results in a one percent point increase in rebound rate for the team. This was true across all positions ("Diminishing Returns and the Value of

Offensive and Defensive Rebounding," 2008). These findings support the analysis of rebounding value based on comparisons to positional rebounding rates.

v. Defensive impact by individuals. The final aspect of the WS model is defensive WS. However, even given the large increases in available data to analyze the NBA, most insights have been focused on offense. A group at Harvard led by Alexander Franks and Andrew Miller attempted to begin research into individual defensive value using player tracking data. Their paper attempted to quantify the defensive value of players by considering that defense involves not allowing your individual matchup to score while also helping teammates defend their matchups. Defensive assignments often switch throughout a possession so allowing your matchup to score could be a result of poor play by the previous defender which permitted an easy shot following a switch in a pick and roll situation. To account for this, Franks and Miller created three separate measures to assess the defensive result versus the expectation of the possession (while accounting for the shooting accuracy of the offensive players).

These three metrics were the original defender, shot defender, and fractional defender. The original defender metric is made up the average number of attempts and points scored against the original defender on the play; the shot metric is computed using the defender at the moment the shot is taken; and, the fractional metric equally awarded the results based on the amount of time the scorer was defended by each player on a given possession. The fractional method was used when ranking players in their results because it can be shown that the original player's matchup will shoot approximately 20% of the time (as you would expect with 5 players sharing the court), but different positions will defend the actual shot at varying rates. Roy Hibbert was used as an example as he defended the shot 45% of the time but his original man was only the shooter on between 20 and 25% of the shots (Franks, 2015). Thus, more than just the final or original defender goes into defending a shot.

The insights added to shooting, passing, rebounding, and defense by these papers will help shape how CMP will assign individual player value given each possession outcome. The findings from these papers provide empirical support or initial findings that will contribute to the CMP model for possession expectations.

III. Data

Stats Inc. partners with the NBA to produce unique movement based statistics for NBA teams and the public. Each NBA arena has had SportVU cameras installed that track every player on the court and where the ball is at every moment during the game. For this paper, Stats Inc. provided a shot chart that included every "player possession" of the 2015-2016 season that ended in a shot or turnover. The 2015-2016 season data is used because that is the first season with reliable shot clock data for the possessions. A "player possession" constitutes the amount of time a player possesses the ball (a team possession is made up of many player possessions).

Combining the player possessions that end in a shot or turnover results in a data set with explanatory variables for every NBA possession outcome in addition to any shot that occurs before an offensive rebound.

Among the variables within this data set are the time on the shot clock at the time a shot is taken, the time remaining in the game (measured in seconds), the result of the possession (make, miss, or turnover), the name of the shooting player, the name of the closest defender, whether or not there was a potential assist, the score of the game, and the potential assister. Unfortunately, whenever a play results in free throws the shot clock is reset so the possession for the shot clock is 24. As a result, the shot clock for foul shot attempts is estimated using the max of 24 and 26 minus the difference in time remaining from the previous possession to the current possession. Twenty six is used as an attempt to account for the time required for the shot clock to be reset.

Unfortunately, even after estimating these changes there are still 88 entries that can be determined to have false data readings. This is because these entries have points scored on a possession either below 0, above 4, or a negative shot clock estimate. These entries have been removed and the players whose shots and shot defenses were affected are displayed in the bottom row of Figure 3. Fortunately, there does not appear to be a pattern causing the errors as no player loses more than 3 possessions as a result of these errors. As a result, removing these entries should not affect our final results.

Figure 3 also shows a few important summary statistics from the remaining data. One particularly interesting chart is the summary of "CLOCK_EST." "CLOCK_EST" is the amount of time remaining on the shot clock at the time of the shot. Interestingly, we find that the first quartile is the most dispersed as it extends for over 8 seconds, while the interquartile range lasts just under 9 seconds. This is important for the hypothesis that high volume scorers add more value when they take shots at the end of the shot clock because Figure 1 clearly shows that the expectation for a possession decreases as the clock gets closer to 0. Therefore, shots made in these tough situations are more valuable to the offense.

Used in conjunction with the data summarized in Figure 3 will be data from the basketball reference database of per game and advanced statistics for the 2015-2016 NBA season. This data includes variables like points, assists, rebounds, PER, WS, and OWS. Additionally, it includes player rebounding and assist rate variables. A player's rebound rate is the percent of available rebounds that a player obtains while he is on the court. Likewise, assist rate is the percent of his teammate's baskets that a player assists on while he is on the court.

Figure 3. Summary Statistics of SportVU Data

```
             PLAYER                   CLOSEST_DEFENDER            POTENTIAL_ASSISTER   CLOCK_EST          MAKE              MISS
 James Harden     :  2353      Draymond Green :  1531                       :139552   Min.   : 0.00   Min.   :0.0000   Min.   :0.0000
 Russell Westbrook:  2051      Damian Lillard :  1426      Rajon Rondo      :  1826   1st Qu.: 8.16   1st Qu.:0.0000   1st Qu.:0.0000
 Stephen Curry    :  2034      James Harden   :  1385      Russell Westbrook:  1763   Median :13.00   Median :0.0000   Median :0.0000
 Paul George      :  1968      Kyle Korver    :  1347      John Wall        :  1737   Mean   :12.80   Mean   :0.4126   Mean   :0.4643
 Damian Lillard   :  1901      Paul Millsap   :  1332      Chris Paul       :  1553   3rd Qu.:17.00   3rd Qu.:1.0000   3rd Qu.:1.0000
 LeBron James     :  1861      Khris Middleton:  1326      James Harden     :  1396   Max.   :24.00   Max.   :1.0000   Max.   :1.0000


       TO                 PTS              PERIOD            POINT_DIF            TIME             POT_AST
 Min.   :0.0000     Min.   :0.0000   Min.   :0.000   Min.   :-54.0000   Min.   :   0.0   Min.   :0.0000
 1st Qu.:0.0000     1st Qu.:0.0000   1st Qu.:1.000   1st Qu.: -7.0000   1st Qu.: 721.3   1st Qu.:0.0000
 Median :0.0000     Median :0.0000   Median :2.000   Median :  0.0000   Median :1444.5   Median :0.0000
 Mean   :0.1231     Mean   :0.9232   Mean   :2.502   Mean   : -0.5965   Mean   :1434.5   Mean   :0.4831
 3rd Qu.:0.0000     3rd Qu.:2.0000   3rd Qu.:3.000   3rd Qu.:  5.0000   3rd Qu.:2163.4   3rd Qu.:1.0000
 Max.   :1.0000     Max.   :4.0000   Max.   :7.000   Max.   : 53.0000   Max.   :2907.8   Max.   :1.0000
```

SUMMARY of REMOVED DATA:

```
        PLAYER                CLOSEST_DEFENDER          POTENTIAL_ASSISTER
 Kawhi Leonard    : 3    No Defender          :15        No Pot Assist    :55
 Kemba Walker     : 3    Brook Lopez          : 2        LaMarcus Aldridge: 3
 Patrick Patterson: 3    DeMar DeRozan        : 2        Alex Len         : 1
 Allen Crabbe     : 2    Dennis Schroder      : 2        Archie Goodwin   : 1
 Carmelo Anthony  : 2    Giannis Antetokounmpo: 2        Chase Budinger   : 1
 Cory Joseph      : 2    Jeremy Lin           : 2        Courtney Lee     : 1
```

- PLAYER: number of times a player appears as the shooter or player committing a turnover in the data
- CLOSEST_DEFENDER: number of times a player appears as the closest defender to the shooter
- POTENTIAL_ASSISTER: number of times a player appears as the potential assister
- CLOCK_EST: estimate of the shot clock at the end of the possession
- MAKE/MISS: whether a shot was made or missed

The final data sets that this paper will utilize are the ESPN salary data for the 2016-2017 season (summary statistics are also included in Figure 4) and the NBA.com list of 2016 free agents. This data includes the age, position, and yearly salary for every player in the NBA. It is for the season following the shot and possession data because the CMP score for the previous year would only have influence on salary for the following year's players that get new contracts. The summary of the ESPN data for the NBA free agents is shown in Figure 4.

The summary shows that there is an even distribution among positions which should allow us to determine if position affects salary. However, there is a very uneven distribution for restricted versus unrestricted free agency that may limit a regression's ability to recognize that unrestricted free agency leads to smaller contracts (by way of the NBA collective bargaining agreement). Thus, it may be necessary to look at only unrestricted free agents because the sample size is too small to look at restricted free agents alone.

Figure 4. Summary Statistics of Salary and Value Data

POS.	AGE	TYPE	DOLLARS	AVG..SALARY
C :19	Min. :21.00		Min. : 98,0431	Min. : 944,691
PF:32	1st Qu.:26.00	RFA: 17	1st Qu.: 2,229,953	1st Qu.: 1,551,659
PG:22	Median :29.00	UFA:110	Median : 12,300,000	Median : 6,125,000
SF:20	Mean :29.28		Mean : 28,039,775	Mean : 8,729,814
SG:34	3rd Qu.:32.00		3rd Qu.: 42,000,000	3rd Qu.:12,994,144
	Max. :39.00		Max. :152,605,576	Max. :33,285,709

PER	OWS	DWS	OBPM	DBPM	WS
Min. : 2.20	Min. :-0.500	Min. :0.000	Min. :-7.1000	Min. :-8.0000	Min. :-0.300
1st Qu.:11.25	1st Qu.: 0.400	1st Qu.:0.600	1st Qu.:-1.8000	1st Qu.:-1.5500	1st Qu.: 1.250
Median :13.80	Median : 1.100	Median :1.200	Median :-0.9000	Median :-0.4000	Median : 2.300
Mean :13.95	Mean : 1.523	Mean :1.405	Mean :-0.7622	Mean :-0.3323	Mean : 2.928
3rd Qu.:16.40	3rd Qu.: 2.000	3rd Qu.:1.850	3rd Qu.: 0.3500	3rd Qu.: 1.0000	3rd Qu.: 3.900
Max. :28.20	Max. :11.000	Max. :5.500	Max. : 7.0000	Max. : 4.3000	Max. :14.500

IV. Methods

i. Methods for CMP - offense. While basketball may be a game of many moving parts and players, at its core it is simply offense versus defense and I believe modeling this as an equation allows for a more accurate estimation of a player's contribution to his team. This section will detail these exact equations and the methods that will be used to estimate their parts.

The emphasis on efficiency stems from the simplest basketball equation: the points scored in a possession is equal to the offense minus the defense. Break this down a little further and you see that the points scored in a possession is equal to the offensive system plus the offensive ability minus the defensive system plus the defense's ability.

$$\textit{Points scored} = (\textit{Offensive System} + \textit{Offensive Ability}) - (\textit{Defensive System} + \textit{Defensive Ability}) \quad (1)$$

Scaling equation 1 to a whole game will make it clear where a player can add value. Equation 2 explains the outcome of a game:

$$\textit{Points Differential in Game} = (\textit{Possessions} * \textit{Points per Possession}) - (\textit{Opponents Possessions} * \textit{Opponent Points per Possession}) \quad (2)$$

From this equation it is obvious that there are three ways for a team to improve its point differential (and in turn, game outcomes) – improve its points per possession, decrease its opponent's points per possession, or increase the relative amount of possessions a team has. For a player this amounts to adding offensive value, adding defensive value, or adding rebounding value (the only way to have an unequal amount of possessions is to have more rebounds or always take the last shot of a quarter).

If we look back at equation 1 and consolidate the defensive terms then we are left with equation 3.

$$\textit{Points scored} = (\textit{Offensive System} + \textit{Offensive Ability}) - (\textit{Defense}) \quad (3)$$

Fortunately, the points scored in each possession is known. Additionally, we can model (Offensive System – Defense) using the template established by Rao. By modeling points per possession using only factors that are indicative of the situation, a league average estimate for the outcome of an equivalent possession can be created. Like the shot clock variable, it is expected that each of these variables will have their own non-parametric relationship with points per possession. Thus, we will attempt to model the generalized additive model of the form:

$$Points = B_0 + B_1 Potential\ Assist + f_1(shot\ clock) + f_2(point\ differential) + f_3(time\ remaining) + \varepsilon \quad (4a)$$

In this equation, Potential Assist is a dummy variable indicating whether or not the final action of the possession was preceded by a pass; shot clock is the time on the shot clock when the possession ends; point differential is the point differential of the game at the time the possession ends; and, time remaining is the amount of seconds remaining until the game ends.

Estimating the expected points independent of the offense's ability will allow us to use the residuals of the model as a representation of the offense's ability on any given possession (the offensive ability in this case is essentially the shooter's ability). Keeping the estimation independent of player ability is the reason predictors which may have significant predictive power for the possession outcomes, like the distance of the shot or the distance of the closest defender, are left out of the estimation. However, the potential assist dummy variable is included because the presence of the potential assist is a function of the offense and is a factor that will affect the outcome, while remaining independent of the shooter.

In fact, the value of B_1 can be attributed to the passing player on the given possession. However, from Figure 3 we can see that 48.3 percent of shots result from a potential assist. As a result, we would expect a player on the court to account for some of these potential assists and only the marginal number of potential assists above average can be attributed to the passing player. Unfortunately, we do not have lineup data or statistics on individual potential assist percentages like we do for assist percentages. However, we can use the assist percentages for players as an estimation of their potential assists above average. First, we will calculate a League Average Assist Rate (a weighted average of assist percentages in the league) using equation 5 (where *i* represents every individual player from the 2015-2016 season).

$$LA\_RATE = \sum(AST_i * \ AST\%_i) \ / \ \sum(AST_i) \qquad (5)$$

Using a player's assist percentage divided by the League Average Assist Rate we are left with the player's rate of assist rate compared to league average. To determine a player's passing value we multiply the number of potential assists a player accounts for by B_1 in equation 4a. Then, we can estimate how an average passer would have contributed by dividing his passing value by the player's rate of assist rate. Subtracting this number from a player's passing value results in his season's cumulative marginal value for passing.

Some may argue that this process is not consistent over positions because of course centers would be expected to be below league average. However, given that this metric compares offensive value regardless of position, then this fact simply indicates that centers do not create as much passing value as the average player on the court. By adding the number of shooting points attributed to each player on each possession with each player's cumulative marginal passing value we are left with the total number of offensive points that their ability contributes to the offense.

ii. Methods for CMP – defense. Without data on who players are defending at each second of a possession we cannot create a fractional defensive rating like in the Franks and Miller paper. As a result, we are only able to recreate a similar metric to the shot defense metric by Franks and Miller. Given that what happens post shot attempt (the result) is attributed to the offense, defensive ability must be accumulated prior to the shot. The defense is tasked with making the offense work as hard as possible to score. As a result, they can be evaluated based on the shot result they force. Specifically we could compare the difficulty of the shot forced with the average expectation of the possession (equation 4a). Unfortunately, our data does not include variables like distance of the shooter or number of dribbles that can explain the shot difficulty. Fortunately, the shot result is likely to be correlated with the actual difficulty (especially when we assume the offensive ability of the opponent evens out over the course of a season). Thus, a first defensive metric can be created that is simply the residuals of equation 4a for the defenders.

However, we have already established that these residuals represent the value of the offensive player. Thus, we can try to estimate the effect of the defensive system and attribute this value to the players under the assumption that defense (more than offense) is a sum of the whole lineup rather than an individual effort. Using a restricted version of equation 4a to estimate the expectation of the possession, defensive value can be assigned to the difference between the expectation for a league average possession and the expectation for the actual possession.

$$Points = B_0 + f_1(point\ differential) + f_2(time\ remaining) + \varepsilon \qquad (4b)$$

Thus, defensive ability on a possession is equal to the difference in equation 4a and equation 4b. In other words, defensive effectiveness is being measured by the ability of the defense to force offenses into lower expectation possessions. Only point differential and time remaining are used in this restricted prediction because point differential and time remaining are independent of the defense but are still likely to affect the outcome of the league average possession. In the same manner that the awarded points for each offensive player were summed, it is possible to sum the points awarded to each defender (points when they are the primary defender on a play) to create the cumulative possession value added defensively.

iii. CMP methods – possessions. The final way players add value is influencing the relative number of possessions his team has in a game. Because our defensive metric only accounts for shot defense we can include forced turnovers in the estimate for "possessions added." Additionally, we saw from the rebounding literature that rebounding is best analyzed based on positional comparisons in which a one percentage point increase in rebounding rate at a position increases the team's rebounding rate by one percentage point. As a result, we can assign rebounding value as a player's rebounding rate divided by the league average rebounding rate for his position times the total number of opportunities he has:

$$Rebounds\ Added = (Reb\% - \overline{Reb\%_{pos}}) * Opportunities \qquad (6)$$

This leaves us with the estimation of Possessions Added:

$$Possessions\ Added = Rebounds\ Added + Turnovers\ Forced^{1} \qquad (7)$$

1. Turnovers forced includes plays where the defender is the primary defender on a turnover and blocks when the defensive residual is not a metric used to evaluate defense.

$$Possessions\ Points\ Added = Possession\ Added * Average\ Points\ per\ Possession \quad (8)$$

Equation 8 is simply multiplying the estimated number of possessions a player added by the expected points scored on a possession. This creates an estimate for the number of points added by these possessions. As a result, we can combine these points added with the defensive points added to create a defensive points added estimate consistent with Win Shares separation of offensive wins and defensive wins. Finally, we can combine the offensive and defensive points added to create the statistic Cumulative Marginal Points (equation 9). Turnovers forced is not standardized to the league average or a positional average because causing a defensive turnover is considered an equivalent to offensive turnovers. Offensive turnovers are considered a possession "used;" similarly, defensive turnovers forced are considered a possession "used" on defense.

$$CMP = CMPO + CMPD \quad (9)$$

iv. Methods for evaluating volume v. efficiency. Given players' calculated CMP scores we can evaluate the metric by comparing these scores with their respective scores for current statistics like Win Shares and PER. Since all three metrics will be on different scales, we can standardize the scores and compare each player's standard deviation from the mean for each metric. Given these rankings, we can determine the players with the largest discrepancies between their PER and WS ratings and their CMP rating. We can then look to see if there is a pattern among these players. For instance, players who attempt a certain threshold of shots may be more likely to have a discrepancy in ratings if it is true that high volume scorers are not properly evaluated by current metrics. Additionally, we can do the same analysis comparing the offensive component of CMP and Offensive Win Shares because WS is made up of, and reported as, an offensive and defensive component.

Furthermore, we can begin to look at possible value of CMP in the league's landscape (note: because this analysis is an evaluation of CMP it will appear in the discussion section). We can run three separate regressions on salaries to determine how teams value players based on WS and PER and if they are already evaluating players based on the factors that lead to CMP. These three regressions can be represented by equation 10.

$$Salary = B_0 + B_1 Experience + B_2 Age + B_{3-6} Position + B_7 metric_i \quad (10)^2$$

Equation 10 predicts salaries from the 2016-2017 season using only players that signed new contracts following the 2015-2016 season in order to make sure that the metrics are actually a part of the salary decisions. Each of the three regressions will include one of the standardized metrics (WS, PER, CMP). The results of these regressions should indicate how these three metrics currently affect salary decisions. Although CMP is not a used metric that teams could base their decisions on, it is possible that teams were already valuing the efficiency-volume interaction that CMP attempts to account for.

V. Results

i. Offensive Model. The first step in developing the CMPO (the entire offensive portion of CMP) for the 2015-2016 season is developing the model to estimate each possession. In section IV we hypothesize that the optimal model will be a Generalized Additive Model of the form from equation 4a. We will also test combinations of these variables for linear regression, polynomial regression, and spline regression. We will use the cross validation error and AIC of

2. Age^2 is not used in this regression because unrestricted free agents have been removed from this analysis. This eliminates the younger player's with rookie contracts. Additionally, the NBA CBA creates a system where as a player ages his maximum per year salary increases. Without the unrestricted free agents there should theoretically be no exponential growth or slowing of salary increases caused by age.

the models to determine which model is optimal for predicting possession outcomes. The results of these models are shown in Figure 5.

Figure 5. **Model Selection**

	Model	CV Error	AIC
1	GAM1	1.24	823987
2	GAM2	1.21	818436
3	GAM3	1.21	818414
4	LM1	1.27	824183
5	LM2	1.22	818752
6	LM3	1.22	818755
7	POLY1	1.24	823808
8	POLY2	1.21	818273
9	POLY3	1.21	818266
10	SPLINE1	1.24	823995
11	SPLINE2	1.21	818438
12	SPLINE3	1.21	818420

In each of the models listed in Figure 5 the progression from 1 to 3 involved first only using the shot clock at the time the possession ended as a predictor. The second model also included the time remaining in the game and a binary variable if there was a potential assist. The third model also included an interaction between the time remaining in the game and the point differential. Model 9 (a polynomial regression) had the lowest Aikake Information Criteria and no discernable difference in cross validation error. Figure 6 shows the summary statistics for this optimal multiple polynomial regression model.

One coefficient to take note of is the coefficient for PA. This implies that a potential assist increases the expected points on a possession by .319 points. Compared to the literature this seems high because the literature showed only an 8% increase in field goal percentage points from a potential assist. However, this eight percentage point increase is a 19% increase in the field goal percentage. Thus, when accounting for both 2 point attempts and three point attempts, the .319 points added from this model are slightly lower than .38 points added (if every field goal attempt was a 2) as is the case in the literature.

Figure 6.

Model Results

	Dependent variable:
	PTS
CLOCK_EST	-0.180**
CLOCK_EST^2	0.247***
CLOCK_EST^3	-0.132***
CLOCK_EST^4	0.037***
CLOCK_EST^5	-0.006***
CLOCK_EST^6	0.001***
CLOCK_EST^7	-0.00004***
CLOCK_EST^8	1.33 e-06***
CLOCK_EST^9	-2.73 e-08***
CLOCK_EST^10	2.38 e-10***
PA	0.319***
PD	-0.001***
PD^2	3.87 e-06
TIME	-0.00004
TIME^2	3.19 e-08
TIME^3	-8.37 e-12*
PD*TIME	6.20 e-07**
PD*TIME^2	1.14 e-11
Constant	0.733***
Observations	269,963
Akaike Inf. Crit.	818,266.000
Note:	*p**p***p<0.01

ii. Shooting results. Using the model's predictions and summing the residuals we produced the Cumulative Marginal Possessions value of each player's "used possessions" (shot attempts or turnovers). The top 5 and bottom 5 players are displayed in Figure 7.

Figure 7. **CMP Shooting Results**

TOP	PLAYER	CMP Shooting	BOTTOM	PLAYER	CMP Shooting
1	Stephen Curry	508.436	429	Emmanuel Mudiay	-183.3583
2	Kevin Durant	292.9453	428	Matt Barnes	-140.8966
3	James Harden	284.1177	427	Alex Len	-140.1566
4	DeMar DeRozan	267.7067	426	Corey Brewer	-121.0809
5	LeBron James	257.613	425	Jared Sullinger	-115.6918

What may be surprising at first is that the worst players are those who play lots of minutes and may even start. However, this actually makes sense given the model because it is impossible to accumulate negative value without being on the court. Thus, the players with the greatest impact (positive or negative) must be players using large shares of the possessions. Presumably, coaches still make the proper decisions about who to play. As a result, this metric cannot necessarily be viewed as an estimation of talent so much as an estimation of value added to a player's team. While this statistic indicates that an average player would score 183 more points than Emmanuel Mudiay had he used the same possessions that Mudiay did, it could still be possible that his playing time was more valuable than someone on their bench that had a better CMP Shooting (closer to 0 but on fewer possessions).

The results of Figure 6 alone could be enough to investigate the efficiency-volume tradeoff. The results correlate to how much value a player's possessions combine to produce. For example, James Harden is often considered a great offensive player, but critics will often point out he shot only 43.9% from the field and 35.9% on threes. These are not efficient percentages and these critics will call him a volume scorer because he is a player that accumulates points due to his volume of shots rather than his efficient percentages. However, the results of CMP Shooting allow us to determine if his volume is positive or negative. We can see that he is actually the third most valuable "shooter."

Unfortunately, Figure 6 is not sufficient for this paper's investigation of the efficiency-volume tradeoff because current efficiency metrics incorporate more than simply shooting. Thus, we must continue to look at the more holistic version of the metric, CMP, in order to compare apples and apples.

iii. Passing results. While Figure 7 represents results for the Cumulative Marginal Possessions in which the player was primarily responsible for the outcome, it does not completely cover the offensive portion of CMP. A major improvement from Win Shares to CMP involves the ability to track potential assists in addition to actual assists. Figure 8 displays the results of CMPO after accounting for the value of potential assists.

Figure 8. **CMPO Results**

TOP	PLAYER	CMP Shooting	AST RATING	CMP Pot Ast	CMPO
1	Stephen Curry	508.436	1.634	141.699	650.135
2	Chris Paul	216.152	2.556	301.408	517.560
3	James Harden	284.118	1.717	185.838	469.956
4	Russell Westbrook	110.300	2.406	328.421	438.721
5	LeBron James	257.613	1.746	168.364	425.977
BOTTOM	PLAYER	CMP Shooting	AST RATING	CMP Pot Ast	CMPO
429	Serge Ibaka	-105.2134	0.1891472	-235.0796	-340.293
428	Andre Drummond	-74.75787	0.2133968	-251.4929	-326.2508
427	Justise Winslow	-88.77453	0.3588947	-170.2856	-259.0601
426	Alex Len	-140.1566	0.4122439	-112.7298	-252.8865
425	Kosta Koufos	-15.01744	0.1600476	-222.5366	-237.554

Figure 8 breaks down how the calculations for CMPO (includes potential assists). The fourth column, "AST RATING," represents the percent of league average a player assists for his teammates. For example, Andre Drummond assists his teammates at a rate 21% of the league average, while Chris Paul assists his teammates at a rate 256% of the league average. Using these rates, the number of points above a league average player that a player contributes through passing to teammates' shots is calculated (as described in Section IV).

CMPO provides a suitable comparison to offensive only metrics like OWS. Theoretically, the efficiency volume tradeoff exists only offensively (although these player types may be correlated to defensive impacts). Consequently, CMPO provides a comparable measurement to begin to draw more concrete conclusions about current metrics' possible bias in evaluating the players and the efficiency-volume tradeoff (see comparison section of results).

The top 5 players in the league for CMPO are the names you would expect to find at the top of an offensive value list. On the other hand, the bottom 5 players may be surprising. There appears to be a trend just from looking at these results, as 4 of the 5 bottom players are essentially centers. Centers are generally considered high efficiency players because they tend to take easy shots and shoot high percentages. However, these centers are predominantly known for their defensive value and rebounding. Furthermore, the recent emphasis on small ball in the NBA suggests that playing without a big man increases offensive productivity because an additional playmaker or shooter is on the court. This may imply that a center's scoring efficiency takes away from teammates' abilities to score efficiently.

iv. Defensive Model and results. Unfortunately, small ball only works if teams are able to maintain defensive efficiency and ball control. Our defensive model will help us look at these big men that are known more for their rebounding and defense. Incorporating the defensive model to CMP provides an apples to apples comparison for PER and WS.

As described in Section IV, we will look at two defensive metrics to observe how players contribute to team defense and to attempt to estimate the difficulty of the shots defenders force.

Additionally, we have incorporated the number of possessions added into CMPD. The results for the best and worst defensively valued players are shown in Figures 9 and 10.

Figure 9 is ranked using CMPD, which uses the method that takes the difference from the result and the possession prediction (results of model in figure 6); while Figure 10 is sorted using CMPD2, a method that takes the difference between the predicted result using shot specific data (results of model in figure 6) and the restricted prediction independent of the actual possession (equation 4b). CMPD is calculated by adding the Shot Dif Est, Rebound Pts Added, and Forced TO Pts Added columns (Rebound Pts Added and Forced TO Pts Added make up Possessions Points Added); while CMPD2 uses Shot Value for Shot Dif Est and includes Blks Pts Added (Rebound Pts Added, Blks Pts Added, and Forced TO Pts Added make up Possessions Points Added). Additionally, the figures include the ratings for each defensive value method to show how players results differ based upon the method used. Both of these methods have their flaws and limits.

Figure 9. **CMPD Results**

Top	CLOSEST DEFENDER	Shot Value	Rebound Pts Added	Forced TO Pts Added	Blks Pts Added	Shot Dif Est	CMPD	CMPD2	CMPD Avg
1	Draymond Green	135.84208	4.0958373	147.94141	-4.276005	133.70449	285.7417	283.6033	284.6725
2	DeAndre Jordan	125.7993	51.7058471	134.41538	79.161324	74.07207	260.1933	391.0819	325.6376
3	Paul Millsap	91.25364	6.5724279	166.78102	35.097294	47.27207	220.6255	299.7044	260.1649
4	Hassan Whiteside	112.71338	49.0729981	87.44815	193.274632	57.8251	194.3462	442.5092	318.4277
5	Kristaps Porzingis	103.94026	-0.5991513	77.76912	54.517153	112.36213	189.5321	235.6274	212.5797
Bottom	**CLOSEST DEFENDER**	**Shot Value**	**Rebound Pts Added**	**Forced TO Pts Added**	**Blks Pts Added**	**Shot Dif Est**	**CMPD**	**CMPD2**	**CMPD Avg**
429	DeMar DeRozan	148.5435	0.5342735	80.32738	-102.95888	-462.4284	-381.5667	126.4463	-127.5602
428	Damian Lillard	192.6242	-2.4873407	112.78603	-91.10252	-390.6908	-280.3921	211.8204	-34.2858
427	Devin Booker	120.6907	-3.7453406	77.00399	-69.23791	-333.9341	-260.6754	124.7114	-67.982
426	Reggie Jackson	170.7497	-3.029577	99.70659	-77.83679	-354.2343	-257.5573	189.5899	-33.9837
425	Jrue Holiday	130.0916	-1.9768341	96.8422	-43.26718	-333.3867	-238.5213	181.6898	-28.4157

Figure 10. **CMPD2 Results**

Top	CLOSEST DEFENDER	Shot Value	Rebound Pts Added	Forced TO Pts Added	Blks Pts Added	Shot Dif Est	CMPD	CMPD2	CMPD Avg
1	Hassan Whiteside	112.71338	49.072998	87.44815	193.274632	57.825102	194.3462	442.5092	318.4277
2	DeAndre Jordan	125.7993	51.705847	134.41538	79.161324	74.072067	260.1933	391.0819	325.6376
3	Andre Drummond	115.19693	83.789182	125.35138	3.453347	-68.88165	140.2589	327.7908	234.0249
4	Karl-Anthony Towns	111.10687	15.664091	138.48886	40.704419	-4.939357	149.2136	305.9642	227.5889
5	Paul Millsap	91.25364	6.572428	166.78102	35.097294	47.272066	220.6255	299.7044	260.1649
Bottom	**CLOSEST DEFENDER**	**Shot Value**	**Rebound Pts Added**	**Forced TO Pts Added**	**Blks Pts Added**	**Shot Dif Est**	**CMPD**	**CMPD2**	**CMPD Avg**
429	Darren Collison	77.590268	-4.553852	117.151483	-234.31346	-78.3279	34.2696	-44.12556	-4.927950
428	Doug McDermott	57.897812	-6.935932	39.678852	-123.97506	-109.2625	-76.5196	-33.33433	54.927002
427	Kevin Martin	44.611529	-1.152793	31.053945	-84.18060	-114.8347	-84.9335	-9.66792	-47.30075
426	Lamar Patterson	11.766558	-0.015892	12.257106	-24.94224	-28.4900	-16.2488	-0.93447	-8.591645
425	Chris Kaman	2.346721	-1.18088	3.841239	-5.793568	-1.2322	1.4281	-0.78649	0.320815

From looking at these two figures it seems that the assumption that big men are valued for their rebounding and defense holds up. However, the presence of guards at the bottom of Figure 9 may be because guards were also the best offensive players. As a result, the players guarding the opposing guards are likely to be scored upon more often than big men simply because they are defending the better players, independent of their actual value. However, if this was the case we would not expect the same player to have a poor rating in Figure 10 because guarding the best players may result in guarding the players shooting at the end of the shot clock and the end of the shot clock is likely to create a net positive rating based on the methods of CMPD2.

Additionally, it could be possible that big men have great ratings in Figure 10 because they are the rim protectors and end up defending many of the end of shot clock drives due to help defense. Thus, averaging the two methods should help differentiate and cancel out players who may be benefiting from the methodology of one of the calculations. These results are displaying in Figure 11 and seem to show that DeAndre Jordan, Draymond Green, and Hassan Whiteside are in fact valuable defenders because they all appear in multiple lists. The opposite can be said for DeRozan, Bogdanovic, and Booker.

Figure 11. **CMPD Average Results**

Top	CLOSEST DEFENDER	Shot Value	Rebound Pts Added	Forced TO Pts Added	Blks Pts Added	Shot Dif Est	CMPD	CMPD2	CMPD Avg
1	DeAndre Jordan	125.7993	51.705847	134.41538	79.161324	74.07207	260.1933	391.0819	325.6376
2	Hassan Whiteside	112.71338	49.072998	87.44815	193.274632	57.8251	194.3462	442.5092	318.4277
3	Draymond Green	135.84208	4.095837	147.94141	-4.276005	133.70449	285.7417	283.6033	284.6725
4	Paul Millsap	91.25364	6.572428	166.78102	35.097294	47.27207	220.6255	299.7044	260.1649
5	Andre Drummond	115.19693	83.789182	125.35138	3.453347	-68.88166	140.2589	327.7908	234.0249
Bottom	**CLOSEST DEFENDER**	**Shot Value**	**Rebound Pts Added**	**Forced TO Pts Added**	**Blks Pts Added**	**Shot Dif Est**	**CMPD**	**CMPD2**	**CMPD Avg**
429	DeMar DeRozan	148.54349	0.534273	80.32738	-102.95888	-462.4284	-381.5667	126.44626	-127.5602
428	Bojan Bogdanovic	109.56204	-5.985438	69.92325	-120.91395	-289.4547	-225.5169	52.58589	-86.4654
427	Devin Booker	120.69068	-3.745340	77.00399	-69.23791	-333.9341	-260.6754	124.71142	-67.982
426	D.J. Augustin	70.46179	-2.230714	40.2085	-61.0871	-207.648	-169.6702	47.35247	-61.15888
425	Ramon Sessions	92.60769	-0.374836	73.23429	-125.50562	-223.2541	-150.3947	39.96152	-55.21658

v. Metric Comparisons. While CMP is a metric designed to improve the deficiencies of PER and WS, it is still important that these metrics have results that look similar to the results of previously accepted metrics. These metrics have been accepted as good estimates for player ability or value; thus, CMP needs to have similar results in order to be taken seriously (especially since it is based on WS). The results for the most valuable and least valuable players based on CMP are shown in Figure 12, along with the top 5 players based on other metrics.

Figure 12. **CMP Results**

TOP	PLAYER	CMPO	CMPD	CMPD2	CMPD Avg	CMP	PER Rk	OWS Rk	WS Rk
1	Stephen Curry	650.1345	111.334104	134.446	122.89003	773.0245	Stephen Curry	Stephen Curry	Stephen Curry
2	Chris Paul	517.5596	-61.806323	157.4069	47.80031	565.3599	Kevin Durant	Kevin Durant	Kevin Durant
3	LeBron James	425.9769	134.51044	127.2978	130.90412	556.881	Boban Marjanovic	James Harden	Russell Westbrook
4	James Harden	469.9556	-9.080447	168.3093	79.61443	549.57	Russell Westbrook	Russell Westbrook	Kawhi Leonard
5	Russell Westbrook	438.7209	13.599111	131.3582	72.47867	511.1196	LeBron James	LeBron James	LeBron James

BOTTOM	PLAYER	CMPO	CMPD	CMPD2	CMPD Avg	CMP
429	Bojan Bogdanovic	-163.0181	-225.5169	52.58589	-86.46548	-249.4836
428	P.J. Hairston	-211.1055	-104.0709	51.32629	-26.37228	-237.4777
427	Dion Waiters	-185.7895	-236.1647	144.47612	-45.84429	-231.6338
426	Trevor Ariza	-224.0789	-133.0416	123.45631	-4.79266	-228.8716
425	Matt Barnes	-227.6539	-121.3831	164.36958	21.49321	-206.1607

The results of the top players for CMP appear similar to but not exactly the same as current metrics. Perhaps the only player surprisingly left out of the top 5 is Kevin Durant as he is second in all three current metrics. On the other hand, second on the CMP list is Chris Paul who is absent from the current metric top 5s. This could suggest that CMP favors players with the ball in their hands as all 5 players could be argued to be point guards and playmakers, while missing players like Durant and Kawhi Leonard (numbers 6 and 14 respectively) can be seen as simply scorers (offensively speaking). This speaks to the impact volume may have on CMP because point guards, like Paul, are likely to be involved in more offensive plays than off ball wing players, like Durant and Leonard.

Figure 13 shows the players with the greatest differential in standard deviation (positive and negative) among metrics. This will help us investigate the actual differences in CMP from PER and WS.

From looking at the three charts in Figure 13 a few things jump out. First, almost all of the most undervalued players by current metrics are point guards, while the overvalued players tend to be big men. Second, Bruno Caboclo could represent a player who does not contribute enough to cause significant negative value to his team (CMP is near 0), but PER values him extremely negatively. There are a few import exceptions to the guards and big men trend, though. Trevor Ariza and Jae Crowder are both overvalued based on Win Shares and are both known as defensive wing players (Ariza is also overvalued by PER). This is significant because it plays into the theory that CMP values ball handling playmakers due to their ability to facilitate teammate scoring and their increased volume of work. This is also significant because the discrepancy in PER for Ariza could be caused by PER being standardized by position while CMP compares every player. OWS would fail to overvalue these defensive players because defense is not included in OWS calculations.

Figure 13. **Metric Comparisons**

PER DIFFERENCE				OWS DIFFERENCE			
UNDER VALUED	PLAYER	CMP.sd	dif.PER	UNDER VALUED	PLAYER	CMPO.sd	dif.OWS
1	Bruno Caboclo	-0.2027714	-3.855272	1	Ish Smith	1.475449	-2.163916
2	Stephen Curry	6.344799	-2.865764	2	John Wall	2.497517	-2.081301
3	James Harden	4.5022932	-2.215347	3	Dennis Schroder	1.152701	-1.941594
4	Jeff Teague	3.0612602	-2.19713	4	Jrue Holiday	2.321903	-1.905688
5	Chris Paul	4.6324893	-2.172498	5	Rajon Rondo	2.342585	-1.876156
OVER VALUED	PLAYER	CMP.sd	dif.PER	OVER VALUED	PLAYER	CMPO.sd	dif.OWS
1	Andre Drummond	-0.7896685	2.307525	1	Tristan Thompson	-1.1969416	3.320394
2	Boban Marjanovic	0.61039961	2.138001	2	Serge Ibaka	-2.6659536	3.232808
3	Enes Kanter	0.02306382	2.013928	3	DeAndre Jordan	-0.6120302	2.936334
4	Trevor Ariza	-1.9163874	1.819156	4	Enes Kanter	-0.5482725	2.872576
5	Shabazz Muhammad	-1.2836672	1.628662	5	Andre Drummond	-2.5399154	2.855705

WS DIFFERENCE			
UNDER VALUED	PLAYER	CMP.sd	dif.WS
1	John Wall	3.027124	-1.966598
2	Jeff Teague	3.06126	-1.966086
3	Ish Smith	1.233717	-1.697752
4	Rajon Rondo	2.212384	-1.532999
5	Dennis Schroder	1.219935	-1.406777
OVER VALUED	PLAYER	CMP.sd	dif.WS
1	Trevor Ariza	-1.9163874	2.872965
2	Andre Drummond	-0.7896685	2.473877
3	Serge Ibaka	-1.4716212	2.462848
4	Jae Crowder	-0.7946575	2.409568
5	Tristan Thompson	-0.150623	2.250621

Finally, an argument could be made that big men are overvalued by OWS compared to CMPO because OWS includes offensive rebounding while CMPO does not. However, this argument does not hold up because when calculating the value from offensive rebounding DeAndre Jordan only produces 18 additional points. This is hardly enough to make a significant difference in his 2.9 standard deviation discrepancy (the other players in the bottom 5 all produce less offensive rebound value than Jordan).

VI. Discussion

i. Limitations of CMP. The results of Section V definitely appear to favor ball dominant players. However, given the nature of the metric maybe this makes sense. The metric is meant to take volume into account, and point guards constantly with the ball in their hands are likely to have some of the greatest usage rates. As a result, the increased usage rate increases the chances of improving their team's results because they must have opportunities to produce positive results. While there may be extremely efficient shooters, if they are not being used as often, they do not have the opportunities to produce a significant positive value. The counter argument is that many wings seem to suffer negative value from potential assists. This could be a result of every player being compared to the average even though certain positions are more likely to produce potential assists. Fortunately, we can recalculate players' CMP using positional averages in a similar manner to PER standardizing by position. This method was not used originally because there is no evidence that positional assist rates have constant returns in the same way that rebounding rates do (Berri, 2006).

An argument can also be made to apply this concept to the standard deviations. Based on the assumption that lineup construction is rather stable, a center could have value over the average center simply because other centers are even worse than the average player. Thus, it may make sense to measure standard deviations by position. However, this concept does not make as much sense for CMP as it does for PER because CMP factors in volume so centers that rarely play would rank as the most valuable if we believe that being a center inherently limits the value a player can contribute. As a result, only an adjustment for position assist rates is shown in Figure 14.

Figure 14. **CMP Passing Alternative Results**

TOP	PLAYER	CMPO Alt	CMP Alt	BOTTOM	PLAYER	CMPO Alt	CMP Alt
1	Stephen Curry	592.6948	715.5848	1	Marcus Smart	-196.67305	-167.8615
2	LeBron James	549.3684	680.2725	2	P.J. Hairston	-114.72284	-141.0951
3	James Harden	572.7061	652.3205	3	Marco Belinelli	-119.70508	-132.447
4	Kevin Durant	463.5857	602.0634	4	Bojan Bogdanovic	-39.93837	-126.4039
5	Chris Paul	467.7403	515.5406	5	Dion Waiters	-80.52477	-126.3691

Following the adjustments we find that most of the top 5 valuable players remains the same, except Kevin Durant (who had been the missing player based on metric comparisons) is now a top 5 player and LeBron James moves up to number 2. At the bottom of the league is where we find the most change. Rather than mostly big men, the entire bottom 5 is comprised of wing players (some of the players that we hypothesized may benefit). Interestingly for the investigation of the efficiency-volume tradeoff, Dion Waiters is the type of player who could be described as a high volume shooter. This adjustment appears to have corrected some of the possible bias against big men and reveals inefficient wing players may be holding offenses back the most (just as efficiency metrics show).

A second criticism of CMP revolves around individual team effects. The model to predict possession outcomes is the same for the entire league. However, in reality this model is likely slightly different for each team. This could have a significant positive or negative impact for players of certain teams. For instance, if one team (perhaps the Rockets) emphasizes playing at a much faster pace than the rest of the league, then the effects from the shot clock portion of the model may not be the same for Rocket's possessions. Because the Rockets shoot earlier in the shot clock, the expectation of a shot taken 4 seconds into the shot clock is likely different than the expectation of a shot taken 4 seconds into the shot clock by a much slower paced team.

On the flip side, some teams could have very poorly constructed offenses or lineups that create different circumstances for a possession. It is possible that a team like the Milwaukee Bucks, which has very few three point shooting threats, has poor floor spacing (compared to the rest of the league), so all of their shots are more difficult due to lineup or offense construction rather than decision making. The results of CMP using a different model for each team are shown in Figure 15 to attempt to investigate the first argument.

Figure 15. **CMP Team Models Alternative Results**

TOP	PLAYER	CMPO Team Alt	CMP Team Alt	BOTTOM	PLAYER	CMPO Team Alt	CMP Team Alt
1	Stephen Curry	404.2829	527.1729	1	Dion Waiters	-76.59101	-122.4353
2	DeAndre Jordan	72.85269	398.4903	2	Marco Belinelli	-108.15426	-120.8962
3	Hassan Whiteside	72.28746	390.7152	3	Emmanuel Mudiay	-151.6835	-114.9083
4	James Harden	301.38454	380.999	4	Stanley Johnson	-88.42145	-114.2678
5	Kevin Durant	240.06241	378.5401	5	P.J. Hairston	-75.66748	-102.0398

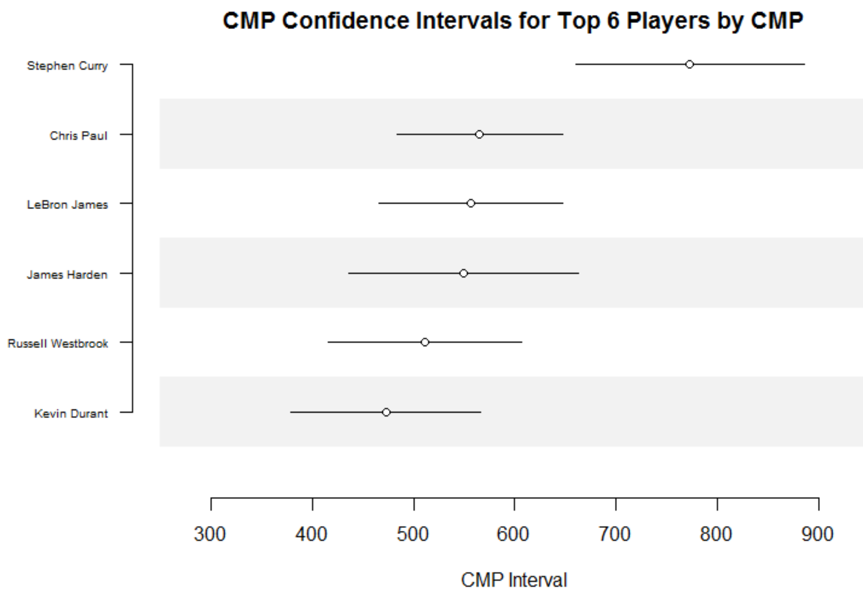
The new results for the bottom 5 players are very similar to the results after adjusting for the positions for potential assists. Like the previous adjustment, big men no longer appear at the bottom of the league. In this case there are actually two big men in the top 5 in the league. This could be a result of their teams constructing below average offenses because they rely on big men. If this were the case, the best big man on that team may end up looking great in comparison to the team's average possessions. Alternatively, the design of this adjustment benefits players who simply play with poor shooting teammates and hurts those who play with better teammates (or even those who increase the performance of their teammates because of their passing). This is why we see only 1 player with CMP over 500 while 6 were above 500 before the adjustments. As a result, this adjustment does not seem as useful as the previous CMP adjustment. However, it could be very useful in evaluating shot selection because some of the players who appeared to be hurting their team by shooting such a high volume were in fact helping their team due to lack of talent/offense around them.

Unfortunately the data used in this paper limits the ability to investigate the claim that team's offenses may force players to take more difficult shots. Further research on this topic utilizing additional tracking data like player distances, dribbles, and touch times could help determine how teams' offenses actually differ (if at all) in creating open shots. Additionally, this research could help alleviate the limitations of CMPD calculations. Actual data on the difficulty of shots forced by defenders could drastically improve CMPD and help illustrate the defensive value of great defensive wing players.

ii. Value of CMP. There are clearly some critiques of CMP as it is currently constructed. However, every metric comes with challenges and limitations. Fortunately, CMP also has further applications that enhance its value in other ways. The first statistical benefit is that CMP is not a point estimate. CMP relies on the results of each individual possession not an entire season's averages or totals. As a result, CMP can be constructed as a confidence interval through bootstraping, while other player metrics are simply a number. By simulating each player's season 1000 times using resampling of each possession we can create a standard deviation for each player's CMPO and CMPD. Assuming these standard deviations are independent we can calculate the standard deviation of CMP and construct a 95% confidence interval for a player's true value.

This is helpful for understanding the consistency and volatility of a player. For example, a player who shoots lots of threes is likely to shoot a lower percentage, but the shots count for more to offset this. However, game to game this player may shoot 6-8 from three and then 0-7 from three. This would be a very good 40% shooter overall, however he did not help his team at all in 50% percent of those games. While this is an extreme example, it illustrates that player value should be viewed as more than just a single estimate. Consequently, constructing a confidence interval for CMP has potential value over other metrics. An example of confidence interval comparisons of the top 5 CMP players is shown in Figure 16. Figure 16 shows that we can say with 95% confidence that Steph Curry's value added was greater than Chris Paul's because there is no overlap in their line segments.

Figure 16.



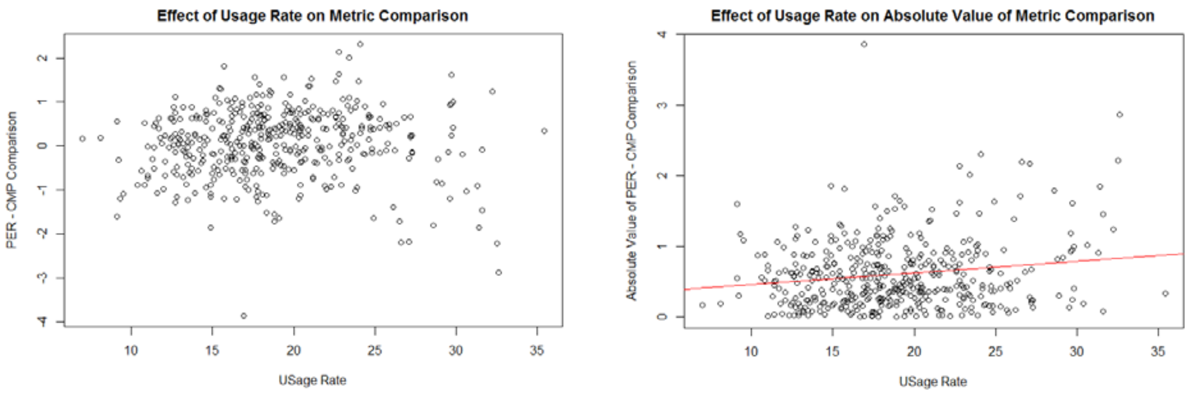
One issue for teams making player acquisitions is when some players play on a bad team it can be hard to identify whether their numbers are a result of being the best bad players or actually being talented. This is especially difficult when measuring counting stats like points and rebounds because someone has to score and someone has to get rebounds, regardless of how bad the team is. Because CMP involves volume as well as the efficiency of each possession, CMP is capable of comparing how a league average player would have fared given that player's circumstances. This comparison should provide an idea of how players on bad teams would have performed in the shoes of players on other teams.

Another application of CMP could actually be in evaluating coaches. While CMP is designed to evaluate players, the assumption that the situation at the time of the shot is indicative of the offensive system minus the defensive system could allow CMP to actually calculate how well coaches' offenses defeat the defense. That is not the design of this paper, but it is a possible, valuable application for the process of calculating CMP.

This paper was, however, designed to address the efficiency volume tradeoff among NBA offensive players. We have seen that current metrics appear to overvalue big men contributions compared to the average player and undervalue point guards. But, what role does volume play in the metric comparison results for CMP? Figure 17 illustrates the relationship between the metric comparisons and usage rate.

From the first graph we can see that there does not appear to be a trend between usage rate and the results of the metric comparisons with PER. This is a good sign because it implies that usage rate should not cause a bias in the results of CMP. However, we do see that the variance for the difference increases as usage rate increases. This implies that greater usage rates lead to less predictability of CMP from PER. Thus, CMP is affected by volume more than PER (as it should be based on the design). This is backed up by the graph on the right which shows a small upward trend in the relationship between usage rate and the absolute difference in PER and CMP ratings.

Figure 17. **Usage Rate and PER Metric Comparisons**



Finally, while there are potentially valuable applications of CMP beyond its designed evaluation of players, CMP must first prove to solve an inefficiency before it can begin to gain value in front offices. In other words, if teams already understand the information that CMP produces without using CMP, then there is no need to adopt CMP. To determine if this is the case we will attempt to predict salaries of free agents using current metrics and CMP. If CMP predicts salary as well as or better than current metrics it can be concluded that front offices already understand the information added by CMP. If CMP performs worse than the other metrics, then there is a potential market inefficiency that CMP may be able to help solve. This information will not prove that CMP solves a market inefficiency, just that the information provided by CMP is not already involved in salary decisions.

We will run three linear regressions on a training set of the data for players who were unrestricted free agents (there are not enough observations of the players who are unrestricted free agents) following the 2015-16 season. Each model will predict average annual salary and will contain control variables of the player's age and position. However, the first two models will include PER and WS, respectively, while the third model will use CMP. After obtaining results we will test the predictions on a test set of the free agents to determine the differences in mean squared error for the average annual value of a player's contract. From Figure 18 we can see that PER and WS are far more predictive than CMP. This suggests that if there is an inefficiency caused by the efficiency-volume tradeoff, then CMP can potentially capitalize on it. However, future analysis of salaries and CMP is still necessary to determine if an inefficiency exists and how well CMP is predictive of this value.

Figure 18.

Salary Prediction Power

Model	Adjusted R-Squared	Test MSE*
PER Model	0.2982243	2,527,289,531
WS Model	0.5004409	2,079,055,840
CMP Model	0.1973856	3,325,402,319

*MSE measured in millions

VII. Conclusion

CMP as a metric attempts to combine shooting efficiency with volume. This combination successfully creates a metric where players that rarely play represent average value. This is different from current metrics because PER only factors efficiency and Win Shares is calculated in a way that last year only 21 of 429 (4.9%) observed players had a negative WS. Thus, even without assessing CMP's ability to evaluate high volume shooters, CMP provides an easier way to determine below average players.

The results of the relationship between usage rate and the metric comparisons suggests that PER and WS could be evaluating high volume players improperly. However, much of this may be driven by the value placed on playmaking that hurts big men and may overvalue the contributions of point guards. Even given the possibility of overvaluing point guards and undervaluing big men, CMP has value in comparing players of the same position. While it may be expected for point guards to have higher scores, you can still differentiate which point guards are best or which shooting guards provide the most value. Additionally, the alternative CMP in which passing is position based provides a solution to the overvaluing of point guards. This approach appears to overvalue wing players who are given ball handling responsibilities like Lebron James and Giannis Antetokounmpo. However, given the indication that big men have limited offensive ceilings, any player that can contribute as a playmaker and rebounder seems extremely valuable because the ability to maintain rebounding without a big man on the floor increases the offensive potential of a lineup.

In conclusion, CMP offers potentially valuable information for teams making player and coaching decisions. The complications of CMP in completely evaluating the efficiency volume bias either arise from issues of scaling CMP to include all aspects of the game that similar metrics do or from limitations in complete tracking data. Still, the foundation established by evaluating each possession a player uses clearly indicates big men are overvalued by current efficiency metrics. While centers are likely to shoot high percentages, when viewed in terms of possession predictions, big men often fail to create above league average value.

VIII. References

Berri, David, Martin Schmidt, and Stacey Brook. *The wages of wins: Taking measure of the many myths in modern sport*. Stanford University Press, 2006.

"Calculating PER." Basketball-Reference.com. Sports Reference, n.d. Web. 15 Dec. 2016.

"Diminishing Returns and the Value of Offensive and Defensive Rebounds." Count The Basket. N.p., 5 Feb. 2008. Web. 15 Dec. 2016.

Franks, Alexander, et al. "Counterpoints: Advanced defensive metrics for nba basketball." *9th Annual MIT Sloan Sports Analytics Conference, Boston, MA*. 2015.

"Game Charting Insights: The Value of Good Pass." 82games. N.p., 2006. Web. 15 Dec. 2016.

Goldman, Matt, and Justin M. Rao. "Allocative and dynamic efficiency in nba decision making." *In Proceedings of the MIT Sloan Sports Analytics Conference*. 2011.

Lyons, Robert, Dr., E. Newton Jackson, Dr., and Aaron Livingston, Dr. "Determinants of NBA Player Salaries." The Sport Journal. United States Sports Academy, 29 May 2015. Web. 15 Dec. 2016.

Maheswaran, Rajiv, et al. "Deconstructing the rebound with optical tracking data." *Proc. 6th Annual MIT Sloan Sports Analytics Conference*. 2012.

Morris, Benjamin. "Stephen Curry Is The Revolution." FiveThirtyEight. ESPN, 03 Dec. 2015. Web. 15 Dec. 2016.

"NBA Win Shares." Basketball-Reference.com. Sports Reference, n.d. Web. 15 Dec. 2016.

Oliver, Dean. Basketball on Paper: Rules and Tools for Performance Analysis. Washington, D.C.: Brassey's, 2004. Print.

"WAR Explained" Baseball-Reference.com. Sports Reference, n.d. Web. 20 Mar. 2017.

"2015-16 NBA Player Stats: Per Game". basketball-reference.com. Retrieved October 26, 2016.

"2015-16 NBA Player Stats: Advanced". basketball-reference.com. Retrieved October 26, 2016.

Appendix 1: Helpful Definitions

Assist – Pass that directly leads to a teammate scoring

Assist Rate – Percentage of a player's teammate's baskets a player assists on while on the court

CMP – (Cumulate Marginal Possessions) an estimate of the number of points added by a player compared to the league average player over the course of a season

DBPM – (Defensive Box Plus Minus) an estimate of the point differential per 100 defensive possessions a player is on the court *not discussed in this paper

DWS – (Defensive Win Shares) estimate of the number of wins a player contributes though defense

Effective Field Goal % - weighted average of the percent of baskets made by a player. 3 point attempts are weighted 1.5 times greater than 2 point attempts

Field Goal % - percent of field goal attempts a player makes

OBPM – (Offensive Box Plus Minus) an estimate of the point differential per 100 offensive possessions a player is on the court *not discussed in this paper

OWS – (Offensive Win Shares) estimate of the number of wins a player contributes though offense

PER – (Player Efficiency Rating) an estimate of the net positive or negative contributions a player makes on a per minute basis. Standardized by position to 15.

Potential Assist – a pass that directly precedes the end of a possession.

PPG – points per game

Rebound Rate – percent of available rebounds a player rebounds while on the court

Shot Clock – each NBA possession requires the possession to end or the ball to hit the rim within 24 seconds. The 24 seconds are counted down on the shot clock.

SportVU Data – Player and ball tracking data

True Shooting % - weighted average of the percent of baskets made by a player. Differs from effective field goal % because free throws are also included.

Usage Rate – percentage of a team's possessions that a player records a shot attempt, turnover, or assist while he is on the court.

WAR – (Wins Above Replacement) baseball metric estimating the wins added above a replacement level player

WS – (Win Shares) basketball metric that estimates the number of a wins a play produces in a season