# Determining NBA Free Agent Salary from Player Performance 

Josh Rosen, Duke University

April 15, 2016
Thesis Advisor: Peter Arcidiacono
Seminar Advisor: Kent Kimbrough


#### Abstract

NBA teams have the opportunity each offseason to sign free agents to alter their rosters. Using only regular season per game statistics, I examine the best method of calculating a player's appropriate salary value based upon his contribution to a team's regular season win percentage. I first determine which statistics most accurately predict team regular season win percentage, and then use regression analysis to predict the values of these metrics for individual players. Finally, relying upon predicted statistics, I assign salary values to free agents for their upcoming season on specific teams. My results advise teams to rely heavily on Player Impact Estimate ("PIE") when predicting their teams' win percentage, and to seek players whose appropriate salaries would be significantly more than their actual season-long salaries if the free agents were to sign.


Duke University, Durham, North Carolina

## Table of Contents

I. Introduction ..... 4
II. Literature Review ..... 9
III. Theoretical Framework ..... 12
IV. Data ..... 14
V. Empirical Specification ..... 21
VI. Team Findings ..... 26
A. Manual Elimination ..... 27
B. Principal Components ..... 41
VII. Player Findings ..... 45
A. Equation 4 ..... 45
B. Equation 6 ..... 50
VIII. Assigning Dollar Values ..... 52
IX. Conclusion ..... 63
References ..... 68
Glossary ..... 70

## Acknowledgments

I would like to thank my thesis advisor Professor Peter Arcidiacono for answering my countless questions when I seemed lost and confused. Additionally, I would like to thank Professor Kent Kimbrough, my seminar advisor, for consistently providing me with comments on my writing and empirical framework, as well as sharing my passion for understanding the game of basketball. Thank you to my peers, especially my roommate, who put up with me thinking out loud about this work for the past few months. Lastly, thank you to my parents for always being there for me along the way.

## I. Introduction

Every offseason, National Basketball Association ("NBA") front offices have the opportunity to reconstruct their teams with free agents. Free agents are players not under contract with a team, and are thus eligible to sign a contract with any team. Players become free agents in two ways - going undrafted out of college or finishing their contract with another team. There are also two types of free agents - unrestricted and restricted free agents. In restricted free agency, the player's original team can retain the free agent via matching an offer that the free agent signs with a different team, while unrestricted free agents can sign with any team regardless of their original team's wishes (Coon, 2015).

NBA free agency, or the period in which players officially become either unrestricted or restricted free agents, starts on July 1 of each year. Teams immediately begin negotiating contracts with targeted free agents, hoping to reach a salary agreed upon by both parties. Players and teams, however, cannot sign negotiated contracts until July 7, providing the league time to determine the upcoming season's salary cap based upon the previous season's league-generated revenue. After the "July Moratorium", players put pen to paper, ending their free agent status and beginning life under a new contract. By reaching agreements soon after the conclusion of the previous season, players have ample time to assimilate to their new location and team.

Free agency allows teams to alter their rosters without trading players or draft picks, as long as teams have not reached the salary cap - the maximum amount of money a team can spend on its players. The more salary cap available to each team's roster, the more funds each front office has to allocate to unsigned players, effectively claiming these individuals for future seasons. Teams establish objectives for free agency based upon their rosters' needs, last season's team performance, and current salary cap space, among other traits. A playoff team with limited
salary cap space may consider its free agency exploits successful if it signs a veteran at the league minimum salary. Meanwhile, a non-playoff team with large cap space may claim free agency success after landing a 25 year old at the league maximum salary. In total, free agency success from the team's perspective is dependent solely on the team's specific situation.

From the player's perspective, free agency success is dependent on the amount of money received through the new contract, location of the new team, expected playing time, as well as talent level of the new team. Players in the prime of their career may reach for the most money available to them, and sign with a team solely based upon that factor. Players who recently purchased a house may decide to resign with their previous team, hoping to keep their family in a region considered home. Veteran players in pursuit of statistical milestones may sign wherever their expected playing time is greatest. Meanwhile, veteran players in search of championships often take less money to sign with a competitive team - hoping to receive a championship ring prior to retiring from the game.

After making their goals of free agency apparent, teams and players interested in one another must then negotiate salaries. Using game film, recorded statistics, medical history, etc., teams determine the appropriate prices to sign free agents. Despite the mass quantity of available information on NBA free agents, calculating what to pay players is subjective and team specific. Teams struggle to determine a player's appropriate compensation.

Occasionally, teams are able to underpay players. This can result from two factors: the team discovers talent in a player that other teams do not see, or a player does not understand his true value to the team. Most likely, these instances of underpaying players are combinations of both. Reigning NBA MVP Stephen Curry, for instance, is under a contract paying approximately $\$ 11$ million per season, currently good for fifth highest on his team. Stephen Curry signed his
contract after seasons featuring glimpses of brilliance and injury agony. Right now, in comparison to other superstars in the NBA, Stephen Curry is not receiving the compensation that he deserves, providing the Golden State Warriors with the opportunity to attract quality players with their extra money. However, not all teams have the Warriors' fortune. Often, teams hurt their future by overpaying for player production.

In the summer of 2012, the Toronto Raptors signed ex-New York Knicks shooting guard Landry Fields to a three-year \$20 million contract. The Raptors expected Fields, who shot a dismal 25.6 three-point percentage and 56.2 free throw percentage in 28.7 minutes per game the previous year, to be a strong contributor for the Raptors going forward (ESPN). Thus, the Raptors allocated approximately $11.5 \%$ of their 2012-2013 salary cap to the shooting guard. The Raptors, however, soon realized that paying $11.5 \%$ of their season's salary cap to a back-up shooting guard unable to shoot was a huge mistake. Fields' per game averages from the 20112012 to 2012-2013 season fell from 8.8 points to $4.7,4.2$ rebounds to 4.1 , and 25.6 three-point percentage to 14.3. In addition, in his three-year stint with the Raptors, Fields played in only 107 out of 246 regular season games due to injury (ESPN). Landry Fields' poor performance limited the Raptors' production on the court, while his high salary limited the talent the Raptors could acquire with their 14 remaining roster spots; the Raptors struggled to attract free agents with reduced funds available under the salary cap. Fields' massively overpaid salary prevented the Raptors from becoming major competitors in the Eastern Conference, and proved to be a dark spot on the Raptors' front office's record.

When the Raptors completed their free agency signing with Landry Fields, they believed they improved their basketball team at a reasonable cost. Not until the season began did they realize that they needed a better method of determining his dollar worth using his previous
statistics and attributes. The Raptors probably used one of multiple metrics to estimate Fields' future contribution to winning.

After a season concludes, Win Shares is a popular metric for determining how many team wins a player is responsible for in the regular season. A player's Win Shares, or the team's number of wins that a player deserves credit for obtaining, is calculated using league, team, and player statistics (Basketball-Reference.com). According to Sports Illustrated's Will Laws, the average margin of error since the 1962-1963 regular season between summing a team's collective Win Shares and the real win total is 2.74 wins (Laws, 2015). However, before the season begins, Win Shares is not very helpful. The metric does not take into account changes in roster composition, so players switching teams maintain the same Win Shares value regardless of the new environment. As a result, a free agent's previous Win Shares value is a poor predictor of his future Win Shares value on a new team. Before the season began, the Raptors would have overestimated Fields' Win Shares for the following season.

Current Vice President of Basketball Operations for the Memphis Grizzlies and former ESPN writer John Hollinger uses Estimated Wins Added to value a player's contribution to winning. Like Win Shares, though, Estimated Wins Added is a backwards looking metric that does not take into account a potential change in roster composition around the player. Thus, a player's Estimated Wins Added on a new team is difficult to predict. Once again, the Raptors would have overestimated the number of Estimated Wins Added for Fields on their team.

The Raptors needed a method of predicting future statistical production of free agents based upon previous player statistics, characteristics, and their own roster composition, as well as then calculating the appropriate compensation for this production. I provide NBA teams with said method, and determine how much teams should pay NBA free agents based upon the
player's previous statistics, current age, prior team's last season performance, and new team's roster. The goal is to create a method using commonly recorded metrics to calculate the player's annual salary that best represents his future contribution to team performance, where team performance is measured by regular season win percentage. By using publicly available per game metrics, I provide all NBA teams the opportunity to appropriately pay free agents. Because the salary cap continuously changes based upon league revenues, I determine the amount paid as a proportion of the salary cap. To test the usefulness of my model predicting team success from player production, I compare my model's findings with that of the Win Shares model.

Creating a statistical method for valuing players will provide more complete information to both front offices and free agents. The model will help teams not overpay for player production, and help free agents better understand their market. This will prevent salary disputes between teams and players, like that of Tristan Thompson and the Cleveland Cavaliers. Recently Thompson, a restricted free agent center, pushed to receive a five-year contract worth approximately $\$ 90$ million from the Cleveland Cavaliers. The Cavaliers, though, believed that Thompson's production was comparable to that of Draymond Green of the Golden State Warriors, who received a five-year $\$ 82$ million contract in free agency. The two sides remained \$8-10 million apart in negotiations, and created many distractions for LeBron James and the rest of the Cleveland roster. A proven model for valuing player performance to team success would have prevented such conflicts and distractions from occurring, and would have allowed Thompson and the Cavaliers to quickly reach an agreed upon salary level.

I began my research by gathering team metrics from NBA.com for all teams in the previous ten seasons. The data includes box-score statistics such as team points per game, as well as advanced metrics such as unassisted two-point field goals made. Using these per game
metrics, I can accurately predict team regular season win percentage. By discovering the statistical reasoning behind why teams win, I know which statistics teams should value in free agency. I only use regular season win percentage in characterizing team success due to the smaller sample size of postseason statistics, and thus only use regular season team statistics in the team analysis.

After examining team statistics, I analyze player performance. Using the player's previous statistics, prior team's success, and current age, I predict the player's performance for the following season. Current age proves useful in predicting future player production. Younger players are expected to improve their performance, while the performance of players in their mid 30 s is expected to gradually decline.

Lastly, I assign dollar values to players when joining a specific team. The change in a team's win percentage due to adding a player determines the player's worth to that team. Each team is expected to value free agents differently according to its current roster make-up.

The paper discusses previous literature affecting my research in Section II - specifically how this research draws upon the work of others. The paper then contains the theoretical framework for the research in Section III, followed by the data in Section IV and empirical specification in Section V. Sections VI, VII, and VIII include the analysis of the team, player, and salary findings respectively. Lastly, I finish with concluding remarks on the analysis in Section IX, as well as a discussion of the research's impact going forward.

## II. Literature Review

The idea of using statistics to value players in sports became popular following the release of Michael Lewis' book "Moneyball" about the Oakland Athletics. Lewis describes the

Athletics' change in player valuation from that of the traditional scout's eye-test to statistical analysis (Lewis, 2004). The Oakland Athletics proved that regular season success could result from relying upon statistical analysis, and signed cheaper players than their competitors due to the difference in player valuation techniques.

More NBA teams and independent researchers have explored NBA statistical analysis since the release of "Moneyball", but few teams and researchers have focused on valuing and predicting player performance. Apart from the aforementioned Win Shares and Estimated Wins Added, succinctly valuing player performance is largely ignored. However, Daryl Morey, general manager of the Houston Rockets, encourages his teams to take three pointers. Morey determined that despite the increased difficulty in making longer shots, shooting three pointers is more valuable to team success than taking two pointers (Strauss, 2013). Houston averaged the most three point attempts per game at 32.7 last season, helping them achieve the sixth most points per game in the league at 103.9 (ESPN).

Most NBA statistical analysis has revolved around better understanding on-court performances. Kirk Goldsberry and Eric Weiss of Harvard University analyzed interior defensive analytics through opponent field goal percentages and shot selection when specific big men are on the court (Goldsberry \& Weiss, 2013). Goldsberry and Weiss argue that elite interior defenders are not the best shot blockers, but are those that deter the most shots from being taken near the basket (Goldsberry \& Weiss, 2013). Goldsberry analyzes other defensive metrics with Alexander Franks, Andrew Miller, and Luke Bornn in "Counterpoints: Advanced Defensive Metrics for NBA Basketball"; they record the percentage of shots that each player contests, amount of points scored against defenders, as well as the probability that people will shoot when guarded by particular players (Franks, Miller, Bornn, \& Goldsberry, 2015).

A better understanding of defense is a common theme within NBA front offices as well. Grantland's Zach Lowe detailed the Toronto Raptors' reliance on statistics to determine defensive alignments. The Raptors use SportVU, "a camera system... that records every movement on the floor and spits it back at its front-office keepers as a byzantine series of geometric coordinates", to analyze where their players are on defense (Lowe, 2013). The front office created a code that takes opponents' skill sets into account to determine where players should be on defense, and compares these results to those depicted via SportVU (Lowe, 2013).

Though teams and researchers are finding new ways to understand on-court performance, there is little progress in determining whether these newly analyzed statistics are important insights into team success. Furthermore, there is not a publicly agreed upon method of assigning dollar values to players based upon these statistics, which is what I hope to establish. When creating the model to analyze player salaries based upon future performance, I consider the following work.

From a labor economics perspective, I examine the tournament theory referred to in David J. Berri and R. Todd Jewell's "Wage Inequality and Firm Performance: Professional Basketball's Natural Experiment". Berri and Jewell referred to tournament theory, which suggests that workers at the lower end of the firm's hierarchy receive wages less than their marginal revenue product, while employees at the top of the hierarchy receive wages higher than their marginal revenue product. The deviation in wage from the marginal revenue product serves as motivation to work harder in the firm to reach an overall individual placement goal (Berri \& Jewell, 2004). Because of a player's ease of mobility amongst teams, I do not believe that players need to accept wages lower than their contribution to any team. However, I will return to tournament theory upon conducting my analysis.

## III. Theoretical Framework

Baseball fan Bill James made one of the earliest attempts to predict team wins from statistics. While Major League Baseball organizations insisted on analyzing player and team performance using the "eye-test", James remained dedicated to understanding baseball through a statistical lens. Though initially rebuffed as crazy, James' legacy lives on - not only do teams from other sports rely on statistical analysis, but also James' Pythagorean Expectation formula to predict baseball win percentage is still used today.

$$
\text { Wins }=\frac{\text { Runs Scored }^{2}}{\text { Runs Scored }^{2}+\text { Runs Allowed }^{2}}
$$

Other statisticians have slightly tweaked the value of the squared exponent used in the model, but the general concept remains the same ("Pythagorean Theorem of Baseball"). Statistician Dean Oliver incorporated the Pythagorean Expectation formula into the NBA by changing the runs metric to points. Oliver found that the exponent changed to 14 (Harrel, 2014), while later work by John Hollinger used 16.5 as the shared exponent ("What Is Pythagorean \& Its Sports Betting Impact"). The difference in the exponent is attributed to the inability of predicting wins from only points scored and points against, as every team follows a different strategy.

The Memphis Grizzlies, due to their size and strength on the court, slowed down games in 2014-2015, as evident by their 94.21 PACE for that season - fifth lowest in the league. A team's PACE is the team's number of possessions per 48 minutes. The Grizzlies depended on strong defense in limited possessions to get a win, and were extremely successful in executing this strategy; their win totals for the ' $12-13$, ' $13-$ ' 14 , and ' $14-$ ' 15 seasons were 56,50 , and 55
wins respectively. The Phoenix Suns of 2006-2007, however, won 61 games while executing the opposite strategy. Head Coach Mike D'Antoni encouraged his team to increase the number of possessions that each team had per game. As such, the '06-'07 Suns had a PACE of 98.08, third highest in the league that season, which helped the Suns average 110.2 points per game - nearly four points per game higher than that of the next highest team that season. Both the 2014-2015 Memphis Grizzlies and 2006-2007 Phoenix Suns were extremely successful during the regular season, but the teams attained their success in opposite ways.

Some teams choose not to have a season-long strategy, but rather to alter their strategies based upon the opponent. The New York Knicks isolate Carmelo Anthony on offense when they determine that he has a weak defender on him. Therefore, when the Knicks play teams with a poor small forward defender, their assist totals may decrease from when they play Jimmy Butler (Chicago Bulls), Kawhi Leonard (San Antonio Spurs), or other great wing defenders. Due to the variety of ways to win, there may not be a clear model for predicting regular season win percentage.

Former Oakland Athletics General Manager and current Executive Vice President of Baseball Operations Billy Beane used statistical knowledge of how MLB teams win to determine which players to acquire. He relied upon the statistics of players while on other teams when predicting their production on the Athletics. The ease of using player performance on other teams to predict future performance on new teams, however, may only work for baseball. Baseball is a fairly independent team sport. Though there are nine players in a lineup, only the performances of two players (the pitcher and hitter) dictate whether a ball gets put into play. Furthermore, if the ball is put into play, the number of players that affect the fielding of the ball
is limited. Thus, the performances of teammates do not play a large role on the performance of the individual.

Basketball, though only allowing five players on the court at any given time, is much more of a team sport. Player chemistry - spacing, communication, movement, and team execution - plays a large role in both individual and team success. Kevin Love's bad chemistry with his Cleveland Cavaliers teammates, namely LeBron James, was a major factor in his poor production during his first season with the Cavaliers. The Cavaliers were unable to predict Love's inability to mesh with the team's roster, but could anybody have been able to predict such a result? Because of the strong influence that a team's roster has on player performance, it is difficult to predict how a player will perform after a change in roster composition.

## IV. Data

I use regular season team and player per game statistics published on NBA.com for my data. The data includes team and player per game statistics from the last ten seasons. I obtained team data in order to first determine how teams win regular season games, and obtained player data to then predict player production in future seasons. NBA.com contained 107 team seasonlong statistics, however, not all of these 107 metrics were recorded for individual players as well; only 76 different metrics were recorded for both teams and individual players during the regular season (e.g. points per game, rebounds per game, etc.). As I hope to examine player contributions to team success, only these 76 statistics were analyzed in predicting team regular season win percentage. In Tables 1 and 2 below, each metric's cell contains the per game average ("A"), standard deviation ("S"), and correlation to wins ("C") of a team statistic over the past ten seasons. The metric is highlighted in green to signify a positive correlation to team wins
and red for a negative correlation. Win percentage is the response variable, and is thus recorded in black without a correlation value. Table 1 lists each of the offensive metrics, while Table 2 includes the defensive statistics. In the later Empirical Specification section, offensive metrics will be denoted as X 1 statistics, with defensive statistics as X 2 .

Table 1 - Offensive Metrics

| Win Percentage | X1: Field Goals Made | X1: Field Goals Attempted |
| :---: | :---: | :---: |
| A:50.0; S:15.5 | A:37.1; S:1.6; C:0.4 | A:81.4; S:2.8; C:-0.1 |
| X1: Field Goal Percentage | X1: Three Pointers Made | X1: Three Pointers Attempted |
| A:45.5; S:1.5; C:0.6 | A:6.7; S:1.6; C:0.4 | A:18.8; S:4.0; C:0.3 |
| X1: Three Point Percentage | X1: Free Throws Made | X1: Free Throws Attempted |
| A:35.6; S:1.9; C:0.5 | A:18.3; S:2.1; C:0.1 | A:24.2; S:2.7; C:0.1 |
| X1: Free Throw Percentage | X1: Offensive Rebounds | X1: Assists |
| A:75.6; S:2.9; C:0.1 | A:11.1; S:1.2; C:-0.2 | A:21.5; S:1.7; C:0.4 |
| X1: Blocks Against | X1: Personal Fouls Drawn | X1: Points |
| A:4.8; S:0.7; C:-0.4 | A:20.9; S:1.6; C:0.1 | A:99.1; S:4.4; C:0.4 |
| X1: Plus / Minus | X1: Effective FG\% | X1: Offensive Rating |
| A:0.0; S:4.6; C:1.0 | A:49.6; S:1.9; C:0.7 | A:103.9; S:3.4; C:0.7 |
| X1: Assist Ratio | X1: Offensive Rebound \% | X1: Net Rating |
| A:16.8; S:1.1; C:0.5 | A:26.4; S:2.5; C:-0.0 | A:0.0; S:5.1; C:1.0 |
| X1: Assist \% | X1: Assist to Turnover Ratio | X1: Turnover Ratio |
| A:57.9; S:3.7; C:0.2 | A:1.5; S:0.2; C:0.5 | A:15.1; S;1.1; C:-0.3 |
| X1: True Shooting \% | X1: PACE | X1: Player Impact Estimate |


| A:53.8; S:1.9; C:0.7 | A:94.7; S:2.6; C:-0.1 | A:50.0; S:3.4; C:1.0 |
| :---: | :---: | :---: |
| X1:2 ${ }^{\text {nd }}$ Chance Points | X1: Fast Break Points | X1: Points in the Paint |
| A:13.2; 1.3; C:-0.1 | A:13.0; S:3.0; C:0.1 | A:40.9; S:4.0; C:0.1 |
| X1:\%FGA 2pts | X1: \%FGA 3pts | X1: \%pts 2pts |
| A:77.0; S:4.7; C:-0.3 | A:23.0; S:4.7; C:0.3 | A:61.3; S:3.8; C:-0.3 |
| X1: \%pts 2pt-MR | X1: \%pts 3pts | X1: \%pts Fast Break Points |
| A:20.0; S:4.0; C:-0.2 | A:20.2; S:4.3; C:0.3 | A:13.1; S:2.7; C:0.0 |
| X1:\%pts Free Throws | X1: \%pts off Turnovers | X1: \%pts Points in the Paint |
| A:18.5; S:1.9; C:-0.0 | A:16.5; S:1.4; C:0.0 | A:41.3; S:3.5; C:-0.1 |
| X1: Percent of 2pt Field Goals | X1: Percent of 2pt Field Goals | X1: Percent of 3pt Field Goals |
| Made Assisted | Made Unassisted | Made Assisted |
| A:51.8; S:4.4; C:0.1 | A:48.2; S:4.4; -0.1 | A:85.3; S:4.2; C:0.1 |
| X1: Percent of 3pt Field Goals | X1: Percent of Field Goals | X1: Percent of Field Goals |
| Made Unassisted | Made Assisted | Made Unassisted |
| A:14.7; S:4.2; C:-0.1 | A:57.9; S:3.7; C:0.2 | A:42.1; S:3.7; C:-0.2 |
| X1: Points off Turnover |  |  |
| A:16.3; S:1.6; C:0.2 |  |  |

Table 2 - Defensive Metrics

| X2: Defensive Rebounds | X2: Total Rebounds | X2: Turnovers |
| :---: | :---: | :---: |
| A:30.8; S:1.6; C:0.5 | A:41.9; S:1.8; C:0.3 | A:14.4; S:1.1; C:-0.3 |
| X2: Steals | X2: Blocks | X2: Personal Fouls |


| A:7.4; S:0.9; C:0.2 | A:4.8; S:0.8; C:0.3 | A:20.9; S:1.7; C:-0.2 |
| :---: | :---: | :---: |
| X2: Rebound $\%$ A:50.0; S:1.5; C:0.5 | X2: Opponent Points off <br> Turnovers A:16.3; S:1.4; C:-0.5 | X2: Opponent $2^{\text {nd }}$ Chance <br> Points <br> A:13.2; S:1.0; -0.3 |
| X2: Defensive Rating A:103.9; S:3.3; C:-0.7 | X2: Defensive Rebound \% A:73.5; S:1.9; C:0.4 | X2: Opponent Fast Break <br> Points A:13.0; S:1.7; C:-0.4 |
| X2: Opponent Points in the <br> Paint <br> A:40.9; S:3.3; C:-0.4 | X2: Opponent Less than 5ft <br> FGM <br> A:16.6; S:1.5; C:-0.5 | X2: Opponent Less than 5 ft $\begin{gathered} \text { FGA } \\ \mathrm{A}: 28.2 ; \mathrm{S}: 2.2 ; \mathrm{C}:-0.3 \end{gathered}$ |
| X2: Opponent Less than 5ft $\begin{gathered} \mathrm{FG} \% \\ \mathrm{~A}: 58.9 ; \mathrm{S}: 2.4 ; \mathrm{C}:-0.5 \end{gathered}$ | X2: Opponent 5-9ft FGM <br> A:3.3; S:0.5; C:0.1 | X2: Opponent 5-9ft FGA <br> A:8.3; S:1.1; C:0.2 |
| X2: Opponent 5-9ft FG\% <br> A:39.8; S:2.5; C:-0.3 | X2: Opponent 10-14ft FGM A:2.7; S:0.4; C:0.3 | X2: Opponent 10-14ft FGA <br> A:6.8; S:0.8; C:0.4 |
| X2: Opponent 10-14ft FG\% <br> A:39.2; S:2.3; C:-0.2 | X2: Opponent 15-19ft FGM <br> A:5.8; S:0.8; C:-0.0 | X2: Opponent 15-19ft FGA <br> A:14.4; S:1.8; C:0.1 |
| X2: Opponent 15-19ft FG\% <br> A:40.5; S:1.8; C:-0.4 | X2: Opponent 20-24ft FGM <br> A:5.7; S:0.8; C:-0.1 | X2: Opponent 20-24ft FGA <br> A:15.0; S:2.1; C:0.0 |
| X2: Opponent 20-24ft FG\% <br> A:38.4; S:1.8; C:-0.4 | X2: Opponent 25-29ft FGM A:2.9; S:0.7; C:-0.2 | X2: Opponent 25-29ft FGA <br> A:8.4; S:2.0; C:-0.1 |
| X2: Opponent 25-29ft FG\% <br> A:34.3; S:2.1; C:-0.3 |  |  |

Though most of the signs of the correlations between metrics and wins are expected, it is interesting to note that both offensive rebounds and second chance points are negatively correlated with wins. Often, teams are praised for their ability to rebound on the offensive end, and then take advantage of such rebounds by scoring second chance points. However, teams that offensive rebound and score afterwards are only able to do so if they miss their original shot. Thus, teams have to be inefficient on their original shot to receive offensive rebounds and second chance points, the cause for the negative correlations.

The data is limited to a ten season span due to a changing style of play over time. As shown in Figures 1 and 2 below, NBA teams on average have adjusted their shooting tendencies. In the 2014-2015 season, teams averaged 22.4 three point attempts per game; in the 2005-2006 season, teams averaged just under 16 three point attempts per game. This increase in outside shooting has led teams away from the basket, effectively reducing the number of free throw attempts per game across this span. To avoid further deviations in team per game averages across seasons, and thus a misunderstanding of how teams win games today, the data is limited to the past ten years.

Figure 1 - Team 3-Pointers Per Game


Figure 2 - Team Free Throws Per Game


The team data from the previous ten seasons also includes whether the team is located in the NBA's Eastern or Western Conference. The player data consists of the above offensive and defensive statistics, as well as age, games played, minutes played, previous teams, and previous teams' wins and losses.

The data from NBA.com is extensive due to the inclusion of both offensive and defensive statistics. In addition, NBA.com offers recorded metrics for both teams and players, which is applicable to my study. Other data sources, such as ESPN.go.com or Basketball-Reference.com, feature either limited statistics or inapplicable statistics to both teams and players. Though NBA.com's dataset includes more defensive player statistics than other publicly available datasets, there are still few recorded defensive player statistics compared to both offensive player statistics and team defensive statistics. NBA.com, for instance, displays 15 defensive metrics for teams that they do not record for individual players (e.g. opponent three point percentage).

Unfortunately, the few efforts to record such individual defensive statistics have been done for only one or two seasons, as was done in "The Effect: A New Ensemble of Interior Defense Analytics for the NBA" and "Counterpoints: Advanced Defensive Metrics for NBA Basketball" (Goldsberry \& Weiss, 2013) (Franks, Miller, Bornn, \& Goldsberry, 2015). Thus, this research uses limited individual defensive statistics in the analysis. Without a firm understanding of player defensive contributions, a model may not accurately value player contributions to teams, especially for defensive-oriented players. Also, apart from PACE, there are no statistics recording a player's or team's style of play, which will restrict the ability to predict player performance when joining a new team.

The analysis does not incorporate the value of a player's leadership or locker room presence. Chris Paul is a known leader around the NBA, and is known for getting his teammates to play at a higher level than they would otherwise. Meanwhile, Lance Stephenson has a history of letting his emotions negatively affect his teammates' play. Unfortunately, there is no objective method of valuing a player's leadership or locker room presence, as much of what occurs in the locker room and amongst players stays private. As a result, a leadership or locker room variable cannot be incorporated into a model valuing a player's contributions to the team.

The experience of the team's head coach may influence the team's regular season win percentage. However, there is no source that tracks the coaching experience of each coach during the regular season. As such, I assume that every team's coach is on average equal in skill, and thus insignificant on the outcome of the game.

Lastly, for calculating average salary levels of point guards, shooting guards, small forwards, power forwards, and centers, as will be explained later, I rely upon the salary data
provided by Basketball-Reference.com. In classifying players by position, I rely upon data from ESPN.go.com.

## V. Empirical Specification

The first step of my analysis involves determining the predictors of team regular season win percentage. I regress regular season win percentage on the significant statistics above, and test for a dummy variable indicating the team's conference (Eastern vs. Western Conference). Assuming the season-long averages remained constant for each of the 82 games, multiplying 82 by the predicted win percentage estimates a team's regular season win total. Shown below is an example of the regression equation used to calculate win percentage during the regular season.

$$
\begin{gather*}
\text { Regular Season Win Percentage }=\alpha_{0}+\alpha_{1} \mathrm{X}_{1}+\alpha_{2} \mathrm{X}_{2}+\alpha_{3} \mathrm{X}_{3}  \tag{1}\\
\text { Let } \mathrm{X}_{1}=\text { Team Offensive Statistics; } \mathrm{X}_{2}=\text { Team Defensive Statistics; } \mathrm{X}_{3}=\text { Eastern or Western } \\
\text { Conference }
\end{gather*}
$$

The dummy variable for conference is included to take into account strength of schedule. The Western Conference has been comprised of significantly better teams than the Eastern Conference for the past few years. Thus I assume that Eastern Conference teams, who play other Eastern Conference teams more often than Western Conference teams, are more likely to win regular season games than if they played in the Western Conference against harder competition. If found statistically significant, the effect of this variable would need to be adjusted in future seasons due to roster changes across the NBA. However, barring many dramatic trades across the

NBA, the Western Conference appears to be the more dominant conference in the NBA in the coming seasons.

After predicting regular season win percentage, I then use regression analysis to predict future player performance. The response variables are those metrics found to be significant in predicting team regular season win percentage. An example of a regression model for estimating player performance is shown below, where points per game for the 2015 season is the variable being estimated.

$$
\begin{gathered}
\mathrm{PPG}_{2015}=\alpha_{0}+\alpha_{1} \mathrm{PPG}_{2014}+\alpha_{2} \mathrm{PPG}_{2013}+\alpha_{3} \mathrm{PPG}_{2012} \\
\text { Where } \mathrm{PPG}=\text { Points per game }
\end{gathered}
$$

The past performances incorporated into the regression model extend as far back as the player performed during the past ten seasons - as long as such performances are deemed significant in predicting future player performance. Otherwise, if the regression model indicates that only two seasons worth of data is significant in predicting future player performance, for instance, then only a player's past two seasons are used as predictors in the regression model.

As mentioned, the regression model for team production provides the regular season win percentage based upon the team's season-long per game statistics. By altering the team's per game averages due to roster modifications, I can estimate the change in win percentage stemming from such roster moves. A change in team per game averages and wins due to a roster alteration occurs from free agency signings, which is exactly what I wish to examine. The issue lies, though, in how to alter the team per game averages to truly reflect the statistical value added by a new player. Summing a player's statistics with a team's per game averages does not provide
the team's new per game averages if the player were to join. This can be illustrated by the following example.

Let's assume that the 15-man 2013-2014 Cleveland Cavaliers roster averaged 88 points per game, 36 rebounds per game, and 18 assists per game. If LeBron James were to take the place of one of the 15 players on the roster for the 2014-2015 season, and, assuming his 20142015 statistical production was 25 points per game, 8 rebounds per game, and 7 assists per game, he would not allow the Cavaliers to average 113 points per game, 44 rebounds per game, and 25 assists per game in the 2014-2015 season. This is because LeBron James' minutes on the court would be replacing the starting small forward's minutes from the previous season; the loss of this replaced player's production must be accounted for in predicting the team's future per game averages. Thus, ignoring the loss of the player LeBron James replaces would overestimate the increase in team per game statistics resulting from adding James.

In order to determine the replaced player's value, I assume that LeBron James is replacing an average player at his position on the Cavaliers. In reality, the player that LeBron James replaces could be the $15^{\text {th }}$ man on the roster - somebody who barely played during the regular season. However, this player did not contribute much to the team's performance, and hence his absence from the court for the Cavaliers is irrelevant. It is the on court absence of the small forward that LeBron James now plays instead of that is worth considering.

I classify every player over the past ten seasons as a point guard, shooting guard, small forward, power forward, or center. In the case of James, I calculate the per game statistics of all small forwards on the Cavaliers if the players were to play James' expected number of minutes. Next, I average the new per game statistics of each of the small forwards. By comparing James' statistics with that of the average production of a small forward on the Cavaliers, I can find the
true change in team per game statistics from adding James and losing a player, which is the difference in the per game statistics of James and the average small forward's statistics. This difference can then be added to the team's per game statistics, and the resulting increased win percentage is the true value that James' statistics add to the Cavaliers' winning chances.

When assigning a dollar value to James' contributions to the Cavaliers, I first assume that teams use their full salary cap to acquire a win for each game. This follows the premise that teams want to win as many games as possible, and that if teams could guarantee winning all 82 games during the regular season by efficiently allocating their full salary cap, then teams would do so. Some team owners more than others seek profits from their team in addition to simply team success. But, even these owners seek winning if given the choice to win or not to win, as winning attracts fans and leads to ticket revenue from postseason appearances. In addition, I determine the current league average salary level of a player at each position. Knowing the Cavaliers' percentage point increase in winning due to replacing their team's average producing small forward with LeBron James, I can then calculate the increase in salary that LeBron James deserves over the league average salary level of a player at his position. For example, if the Cavaliers' win percentage jumps 20 percentage points due to the addition of James to their roster, then James deserves to be paid $20 \%$ more of the salary cap than the league average salary of a small forward. In other words, since adding James "guarantees" that the Cavaliers will now win an additional $20 \%$ of the season's games, then assuming that efficiently spending the full salary cap can guarantee winning 82 games, James should be paid $20 \%$ of the salary cap plus the average salary of a small forward. A step-by-step example of calculating LeBron James' value when added to the Cavaliers is shown below, where the simplifying assumption is that the only
team per game statistics deemed significant in predicting team win percentage are points, rebounds, and assists. Each of the numbers shown is created for the purpose of the example.

Example 1 - Calculating LeBron James' Salary

1. Cavs Win Percentage Without LeBron James $=0.50$
2. Cavs Predicted Per Game Team Statistics $=88$ points, 36 rebounds, 18 assists
3. Current Small Forwards on the Cavs Roster: Mike Miller and Shawn Marion
4. LeBron James' Expected Minutes Per Game on the Cavs $=30$ min
5. Mike Miller Predicted Per Game Statistics in 30 min: 12 points, 5 rebounds, 1 assist
6. Shawn Marion Predicted Per Game Statistics in 30 min: 8 points, 3 rebounds, 3 assists
7. Avg Per Game Statistics of Cavs Small Forward $=10$ points, 4 rebounds, 2 assists
8. LeBron James Predicted Per Game Statistics on Cavs $=25$ points, 8 rebounds, 7 assists
9. Added Contribution to the Cavs $=15$ points, 4 rebounds, 5 assists
10. Cavs Win Percentage With LeBron James $=0.70$
11. Predicted Number of Wins Without LeBron James $=41$ Wins
12. Predicted Number of Wins With LeBron James = 57.4 Wins
13. Because LeBron James increases the Cavs' regular season win percentage by 20 percentage points, LeBron James deserves to be paid $20 \%$ more of the salary cap than the league average salary of a small forward.
14. League Average Small Forward Salary $=\$ 6 \mathrm{M}$
15. Team Salary Cap for the 2014-2015 Season $=\$ 63 \mathrm{M}$
16. LeBron James' Appropriate Salary on the Cavs $=\$ 6 \mathrm{M}+(0.20 \mathrm{X} \$ 63 \mathrm{M})=\$ 18.6 \mathrm{M}$

Following this methodology, players receive the average salary per year at their position if their statistical contributions are equal to that of the average player at their position on the team they are joining. Meanwhile, players deserve a salary higher or lower than the league average positional salary based upon whether the player performs better or worse than the team's average positional production.

A player's salary level is calculated from the league average positional salary rather than the team's average positional salary to avoid the effects of a team's previous spending. If Mike Miller and Shawn Marion's salaries were each \$15M in the example above, and LeBron James' proportional increase in salary was due to the team's average positional salary rather than that of the league, then LeBron James' calculated salary would be $\$ 27.6 \mathrm{M}$. As a result, James would be inclined to sign with a team that had already overpaid for players of his position. However, by using the league average salary, James' salary is affected by the league-wide value for players of his position, thus avoiding the influence of a team's history of paying free agents at a specific position.

## VI. Team Findings

To determine a model for predicting team regular season win percentage, I used two different methodologies when performing linear regression. Both methodologies entailed a comprehensive analysis of all of the collected variables for team statistics over the past ten seasons, and effectively selected which variables or components were most important in predicting team win percentage. Furthermore, the selection process used in each methodology removed the multicollinearity that exists among the recorded team metrics.

## A. Manual Elimination

The first methodology, deemed "manual elimination", involved grouping the collected statistics into categories based upon the basketball actions that they describe. After understanding the recorded team metrics, I placed statistics into the following buckets: own shooting, opponent shooting, free throws, rebounding, passing, and miscellaneous. Shown below are the metrics that encompassed each bucket, where the metrics highlighted in red were the selected metrics from each bucket.

Table 3 - Manual Elimination
Own Shooting (29)

| FGM | FGA | FG\% | 3PM |
| :---: | :---: | :---: | :---: |
| 3PA | 3P\% | eFG $\%$ | TS $\%$ |
| \%FGA 2pts | \%FGA 3pts | \%pts 2pts | \%pts 2pt-MR |
| \%pts 3pts | \%pts FBPs | \%pts Off TO | \%pts PITP |
| 2FGM \%AST | 2FGM \%UAST | 3FGM $\% A S T$ | 3FGM \%UAST |
| FGM \%AST | FGM \%UAST | PTS | Off Rtg |
| PTS Off TO | $2^{\text {nd }}$ PTS | FBPs | PITP |
| Opp BLKA |  |  |  |

Opponent Shooting (24)

| Opp Less than 5ft <br> FGM | Opp Less than 5ft <br> FGA | Opp Less than 5ft <br> FG\% | Opp 5-9ft FGM |
| :---: | :---: | :---: | :---: |
| Opp 5-9ft FGA | Opp 5-9ft FG\% | Opp 10-14ft FGM | Opp 10-14ft FGA |
| Opp 10-14ft FG\% | Opp 15-19ft FGM | Opp 15-19ft FGA | Opp 15-19ft FG\% |
| Opp 20-24ft FGM | Opp 20-24ft FGA | Opp 20-24ft FG\% | Opp 25-29ft FGM |
| Opp 25-29ft FGA | Opp 25-29ft FG $\%$ | BLK | Def Rtg |
| Opp PTS Off TO | Opp 2 ${ }^{\text {nd }}$ PTS | Opp FBPs | Opp PITP |

Free Throws (4)

| FTM | FTA | FT\% | \%pts FT |
| :---: | :---: | :---: | :---: |

Rebounding (6)

| OREB | DREB | REB | OREB\% |
| :---: | :---: | :---: | :---: |
| DREB $\%$ | REB $\%$ |  |  |

Passing (7)

| AST | TO | STL | TO Ratio |
| :---: | :---: | :---: | :---: |


| AST\% | AST/TO | AST Ratio |  |
| :---: | :---: | :---: | :---: |

Miscellaneous (6)

| Net Rtg | PF | PFD | $+/-$ |
| :---: | :---: | :---: | :---: |
| PACE | PIE |  |  |

By forming buckets for different metrics, "manual elimination" prevented multicollinearity - where multicollinearity is defined as two explanatory variables in the regression analysis having a correlation greater than 0.8 or less than -0.8 . The own shooting, opponent shooting, free throws, rebounding, passing, and miscellaneous buckets did not have strong positive or negative correlations with one another.

Because the outcome of a basketball game is dependent mostly on shooting statistics (e.g. field goal percentages, field goals attempted, etc.), I chose to select three metrics from each of the shooting buckets, while only selecting one metric from each of the other buckets. I limited the number of metrics selected from the shooting buckets to three in order to prevent the aforementioned multicollinearity from occurring.

From the own shooting bucket I selected FGM (field goals made), eFG\% (effective field goal percentage), and FGM \%AST (percent of field goals made that are assisted). Each of these selections was a subjective decision based upon an understanding of the statistic and prior knowledge of basketball. I chose FGM because the more teams make field goals, the more points that team scores, and thus the better chance the team has of outscoring its opponent. I chose effective field goal percentage because it is a field goal percentage metric that adjusts for the value of three pointers being greater than two pointers; it is more important for teams to make their three pointers than their two pointers due to the difference in shot value. The equation for eFG\% is shown below.

$$
\begin{equation*}
\mathrm{eFG} \%=\frac{(\mathrm{FGM}+0.5 \times 3 \mathrm{PM})}{\mathrm{FGA}} \tag{2}
\end{equation*}
$$

Lastly, I selected FGM \%AST because teams that pass the ball often and well are more likely to find open and easier shots - leading to more points. The Golden State Warriors and San Antonio Spurs are prime examples of offenses that succeed due to unselfish passing.

Unfortunately, NBA.com does not record opponent FGM, opponent eFG\%, or opponent FGM \%AST for individual players. Thus, incorporating these metrics into a team prediction model would prove useless when eventually trying to determine how each player contributes to team win percentage. Using the opponent shooting metrics from Table 3, I selected an array of opponent field goal percentages that describe short, medium, and long range shooting on the court - essentially accomplishing the task of providing an effective field goal percentage for opposing players. I believe that opponent shooting from less than five feet from the basket, 10-14 feet from the basket, and 20-24 feet from the basket accurately summarize the short, medium, and long range shooting that occurs during a game.

I selected FTM (free throws made) from the free throws bucket. Unlike the other free throw based statistics in the group, FTM takes into consideration both the number of times that a team shoots free throws as well as the success of these attempts. The percentage of points that are free throws (\%pts FT) does a similar task to that of FTM, though it is highly dependent on the number of FGM for teams.

I selected DREB (defensive rebounds) from the rebounding bucket. I preferred DREB to REB (total rebounds) and OREB (offensive rebounds) because in order for a team to retrieve OREB, the team must miss its shots. Thus, having a high number of OREB or REB may be a
combination of both strong and poor play for a team. High DREB totals, however, are only positive indicators of a team's performance, as defensive rebounds represent a team forcing the opponent to miss its shots.

I chose the AST/TO (assist to turnover ratio) over other metrics within the passing bucket because AST/TO takes into account both the assist and turnover totals of a team. Teams that total high assist and high turnover numbers or low assist and low turnover numbers do not perform as well as teams with high assist and low turnover counts. Thus, AST/TO is the metric in the passing bucket most telling of offensive success.

Lastly, I selected PF (personal fouls) from the miscellaneous bucket. Though opponent free throws made was not recorded for individual players on NBA.com, PF is a strong indicator of the number of free throws made for the opponent.

I included each of these metrics in the linear regression equation shown below, where regular season win percentage is the response variable. I chose to use linear regression in order to predict the effects of multiple metrics on team win percentage. The results from the regression are shown in Table 4.

$$
\begin{align*}
\text { Winning } \%= & \alpha_{0}+\alpha_{1} \text { FGM }+\alpha_{2} \text { eFG } \%+\alpha_{3} \text { FGM } \% \text { AST }+\alpha_{4} \text { Opp Less than } 5 \mathrm{ft} \mathrm{FG} \% \\
& +\alpha_{5} \text { Opp } 10 \text { to } 14 \mathrm{ft} \text { FG } \%+\alpha_{6} \text { Opp } 20 \text { to } 24 \mathrm{ft} \mathrm{FG} \%+\alpha_{7} \mathrm{FTM}+\alpha_{8} \text { DREB }  \tag{3}\\
& +\alpha_{9} \frac{\mathrm{AST}}{\mathrm{TO}}+\alpha_{10} \mathrm{PF}
\end{align*}
$$

Table 4 - Regression Output from Equation 3

| Variable | Estimate T |  | T-value | Std. Error | P-value | 95\% Lower | $\begin{gathered} \hline 95 \% \text { Upper } \\ \hline 50.36 \end{gathered}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Intercept | -0.80 |  | -0.03 | 26.00 | 0.98 | -51.96 |  |
| FGM | -2.44 |  | -5.36 | 0.45 | 1.67e-07 | -3.34 | -1.54 |
| eFG\% | 4.34 |  | 12.80 | 0.34 | $<2 \mathrm{e}-16$ | 3.67 | 5.01 |
| FGM \% | -0.72 |  | -4.27 | 0.17 | 2.62e-05 | -1.05 | -0.39 |
| AST |  |  |  |  |  |  | -0.47 |
| Opp $<5 \mathrm{ft}$ | -0.86 |  | -4.42 | 0.19 | 1.37e-05 | -1.24 |  |
| FG\% |  |  |  |  |  |  |  |
| Opp 10-14ft | -1.88 |  | -6.50 | 0.29 | $3.53 \mathrm{e}-10$ | -2.44 | -1.31 |
| FG\% |  |  |  |  |  |  |  |
| Opp 20-24ft | -1.04 |  | -4.39 | 0.24 | 1.56e-05 | -1.50 | -0.57 |
| FG\% |  |  |  |  |  |  |  |
| FTM | 1.57 |  | 6.05 | 0.26 | 4.53e-09 | 1.06 | 2.08 |
| DREB | 1.38 |  | 3.79 | 0.36 | 1.83e-04 | 0.66 | 2.10 |
| AST/TO | 39.98 |  | 9.22 | 4.34 | $<2 \mathrm{e}-16$ | 31.45 | 48.52 |
| PF | -0.86 |  | -2.45 | 0.35 | 0.01 | -1.55 | -0.17 |
| $\text { R-Squared }=0.74$ |  |  |  |  |  |  |  |
| $\begin{aligned} & \text { Residual standard error }=7.98 \\ & \mathrm{~N}=300 \end{aligned}$ |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
| Residuals: |  |  |  |  |  |  |  |
| Min | 1Q M | Median | - 3Q | Max |  |  |  |
| -17.69 | -5.77 -0.1 | -0.11 | 5.18 | 22.08 |  |  |  |

P-values bolded are less than 0.05

The r -squared for the model is 0.74 , indicating that approximately $74 \%$ of the variation in win percentage is explained by the independent variables in the regression model. The residuals of the model have a minimum of -17.69 and a maximum of 22.08. Translated into number of games won or lost, the regression model predicts one team to win approximately 15 games more than the team actually won, while the model also predicts one team to win about 18 fewer games than it did.

The $p$-values of each of the variables are less than 0.05 , providing evidence that the coefficients of each of the explanatory variables are significantly different from zero. The signs
of the coefficients for effective field goal percentage, opponent field goal percentage less than 5 feet, opponent field goal percentage 10-14 feet, opponent field goal percentage 20-24 feet, free throws made, defensive rebounds, assist to turnover ratio, and personal fouls all make sense from a basketball perspective. The coefficient estimates for field goals made, field goals made percent assisted, and assist to turnover ratio are interesting to note.

The field goals made variable has a coefficient of -2.44 ( $95 \% \mathrm{CI}:-3.34,-1.54$ ). If the number of field goals made per game for a team increases by one, holding all else constant, then the regular season win percentage for the NBA team is predicted to decrease by 2.44 percentage points. The negative nature of the coefficient for field goals made is certainly counterintuitive from a basketball perspective. As teams make more field goals, they score more points, thus increasing their chances of outscoring the opponent. However, the results of the regression analysis indicate otherwise - that teams decrease their chances of winning with each made field goal. These results illustrated the need for further modifications to the model predicting team regular season win percentage.

The coefficient for percent of field goals made assisted is -0.72 ( $95 \% \mathrm{CI}:-1.05,-0.39$ ). The model indicates that if the percent of field goals made assisted per game increases by one percent, holding all else constant, a team's regular season win percentage will decrease by 0.72 percentage points. The sign of this coefficient is surprising. As previously indicated, teams that often share the ball perform better on the offensive end than those that do not. The Oklahoma City Thunder has demonstrated success doing the opposite; the Thunder rely on few passes and isolation plays to score effectively. I believe that the Thunder, though, are an exception to the rule.

The assist to turnover ratio variable has a coefficient of 39.98 ( $95 \% \mathrm{CI}: 31.45,48.52$ ). If the per game assist to turnover ratio increases by one, holding all else constant, then the team's regular season win percentage will increase by 39.98 percentage points; a one unit increase in the per game assist to turnover ratio can easily be the difference in a team making or not making the postseason. The coefficient for the assist to turnover ratio variable is by far the greatest in magnitude amongst the explanatory variable coefficients. However, due to the nature of what the metric describes, the magnitude of the variable is not all that surprising. In the past ten seasons, only four teams posted an assist to turnover ratio of 2 or greater, every other team posted an assist to turnover ratio between 1 and 2 . Thus, it seems reasonable that a one unit increase in the ratio would yield such a large change in win percentage. Furthermore, the positive sign of the coefficient makes sense from a basketball viewpoint, as teams that increase their assists while diminishing their turnovers are more likely to score than those that do not.

Due to the negative coefficient estimates for both FGM and FGM \%AST, I conducted regression analyses without one of these two variables in the equation. In both instances, the remaining variable in the regression maintained a negative coefficient estimate. As such, I removed both FGM and FGM \%AST from the equation. In addition, to test my belief that DREB is a better indicator of winning than total rebounds, I substituted REB into the regression equation for DREB. Lastly, I included the aforementioned PACE metric into the regression equation with the hope of determining whether speed of play is a predictor of team win percentage, despite my initial thought indicating otherwise. The modified equation is shown below, with the regression output in Table 5.

$$
\begin{align*}
\text { Winning } \%= & \alpha_{0}+\alpha_{1} \text { FTM }+\alpha_{2} \mathrm{eFG} \%+\alpha_{3} \text { Opp Less than } 5 \mathrm{ft} \text { FG } \% \\
& +\alpha_{4} \text { Opp } 10 \text { to } 14 \mathrm{ft} \text { FG } \%+\alpha_{5} \text { Opp } 20 \text { to } 24 \mathrm{ft} \mathrm{FG} \%+\alpha_{6} \text { REB }+\alpha_{7} \frac{\mathrm{AST}}{\mathrm{TO}}  \tag{4}\\
& +\alpha_{8} \text { PACE }+\alpha_{9} \text { PF }
\end{align*}
$$

Table 5 - Regression Output from Equation 4

| Variable | Estimate | T-value | Std. Error | P-value | 95\% Lower | 95\% Upper |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Intercept | -45.28 | -2.00 | 22.62 | 0.05 | -89.81 | -0.75 |
| FTM | 1.79 | 8.30 | 0.22 | 3.89e-15 | 1.37 | 2.22 |
| eFG\% | 4.62 | 19.12 | 0.24 | <2e-16 | 4.15 | 5.10 |
| Opp $<5 \mathrm{ft}$ | -0.55 | -3.43 | 0.16 | 6.89e-04 | -0.87 | -0.24 |
| FG\% |  |  |  |  |  |  |
| Opp 10-14ft | -1.39 | -5.94 | 0.23 | 8.12e-09 | -1.85 | -0.93 |
| FG\% |  |  |  |  |  |  |
| Opp 20-24ft | -0.70 | -3.55 | 0.20 | 4.56e-04 | -1.09 | -0.31 |
| FG\% |  |  |  |  |  |  |
| REB | 2.90 | 11.05 | 0.26 | $<2 \mathrm{e}-16$ | 2.38 | 3.42 |
| AST/TO | 23.03 | 8.73 | 2.64 | $<2 \mathrm{e}-16$ | 17.83 | 28.22 |
| PACE | -2.20 | -12.41 | 0.18 | $<2 \mathrm{e}-16$ | -2.55 | -1.85 |
| PF | -0.60 | -2.13 | 0.28 | 0.03 | -1.15 | -0.04 |
| R -Squared $=0.82$ |  |  |  |  |  |  |
| Residual standard error $=6.61$$\mathrm{~N}=300$ |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| Residuals: |  |  |  |  |  |  |
| Min | 1Q M | n 3Q | Max |  |  |  |
| -15.21 | -4.41 | 4.64 | 19.23 |  |  |  |

P -values bolded are less than 0.05

The r-squared of Equation 4's regression is higher than that of Equation 3's regression ( 0.82 compared to 0.74 ), while the residual standard error is lower ( 6.61 compared to 7.98 ) both indications of a stronger predicting model. The coefficients' signs and magnitudes of the variables included in Equation 3 and Equation 4 (FTM, eFG\%, Opp $<5 \mathrm{ft}$ FG\%, Opp 10-14ft FG\%, Opp 20-24ft FG\%, AST/TO, and PF) remain fairly similar in both regressions. The

AST/TO coefficient has the greatest change in magnitude, with the coefficient estimate changing from 39.98 to 23.03. The AST/TO variable's statistical significance though, like each of the variables included in both Equation 3 and 4, remains the same due to a p-value of less than 0.05 . Furthermore, the added REB and PACE variables are found to be statistically significant due to a p-value of less than 0.05 .

The free throws made per game variable has a coefficient of 1.79 ( $95 \%$ CI: $1.37,2.22$ ) in Equation 4. The coefficient indicates that a team that ended three games below .500 could have finished even on the season averaging only two more free throw makes per game, holding all else constant. The sign of the coefficient seems reasonable from a basketball perspective. Free throws are a high percentage shot that most players prefer to take in comparison to a contested jump shot. Having the opportunity to make a free throw also means that an opposing player committed a foul, which can force the opposition to sit talented players in foul trouble. In addition, the magnitude of the coefficient makes sense. In order for the number of free throws made to greatly affect team win percentage, teams must attempt a significant amount of free throws per game.

Effective field goal percentage has a coefficient of 4.62 ( $95 \% \mathrm{CI}: 4.15,5.10$ ). If the effective field goal percentage of a team increases by one percent, while holding all else constant, the team's regular season win percentage will increase by 4.62 percentage points. The positive sign of the coefficient seems reasonable, as the more efficiently teams shoot on the court, the more likely teams will win games.

Opponent field goal percentage less than 5 feet from the rim has a coefficient of -0.55 ( $95 \% \mathrm{CI}:-0.87,-0.24$ ). If the opponent's field goal percentage less than 5 feet from the rim increases by one percent, holding all else constant, the team's regular season win percentage will
decrease by 0.55 percentage points. Opponent field goal percentage 10-14 feet from the basket has a coefficient of $-1.39(95 \% \mathrm{CI}:-1.85,-0.93)$. If the opponent's field goal percentage $10-14$ feet from the basket increases by one percent, holding all else constant, the team's regular season win percentage will decrease 1.39 percentage points. Opponent field goal percentage 20-24 feet from the hoop has a coefficient of $-0.70(95 \% \mathrm{CI}:-1.09,-0.31)$. If the opponent's field goal percentage 20-24 feet from the hoop increases by one percent, holding all else constant, the team's regular season win percentage will decrease 0.70 percentage points. The magnitudes and signs of these coefficients are reasonable. The more efficiently the opponent shoots from any spot on the floor, the less likely a team is to win the game. It is noteworthy though that the coefficient for opponent field goal percentage from 10-14 feet is greatest in magnitude amongst the three different spots on the floor. I would have assumed that the coefficient for 20-24 feet on the court would have the greatest magnitude or impact on win percentage, as 20-24 feet includes the more valuable three point shot.

The substituted REB variable has a coefficient of 2.90 ( $95 \% \mathrm{CI}: 2.38,3.42$ ), indicating that if total rebounds per game increased by one, holding all constant, then the team's win percentage would increase by 2.90 percentage points. It is interesting that the coefficient and tvalue for total rebounds is larger than that of the DREB variable in Equation 3, illustrating that total rebounds in fact play a larger role in predicting team win percentage than solely defensive rebounds. This is mostly likely because though receiving offensive rebounds only occurs from missed shots, at least the team is making the most of its missed shots rather than going immediately to defend. Due to the statistical significance and importance of total rebounds, I continue to use total rebounds instead of defensive rebounds in the regression analysis.

The PACE variable, despite my initial beliefs, proves to be statistically significant in predicting team regular season win percentage. The PACE variable has a coefficient of -2.20 ( $95 \% \mathrm{CI}:-2.55,-1.85$ ), demonstrating that a one possession increase per 48 minutes, holding all else constant, results in approximately 2 fewer wins in the regular season. Though successful fast breaks are helpful for teams, quick shots are often an indication of poor and inefficient shot selection. Thus, teams that hold the ball to find the best open shot seem to be more likely to win than teams willing to take the first shot available.

Lastly, the personal fouls variable has a coefficient of -0.60 ( $95 \% \mathrm{CI}:-1.15,-0.04$ ). If the number of personal fouls increases by one foul per game, holding all else constant, the team's regular season win percentage will decrease 0.60 percentage points. The magnitude of the coefficient is fairly small, though statistically significant. This makes sense, as an increase in only one personal foul per game is not likely to make a big impact on a team's win chances. The negative sign of the variable's coefficient is intuitive, as personal fouls hurt a team's chances of winning a game. Personal fouls prevent players from staying on the court, as well as often allow the opposition to take free throws.

In addition to using Equation 4 to predict team win percentage, I conducted a regression analysis using net rating as the only explanatory variable to predict team win percentage. Net rating is the difference in a team's offensive and defensive rating; offensive rating is the number of points scored for a team in 100 possessions, while defensive rating is the number of points allowed by a team in 100 possessions. When calculating net ratings for individual players, only the points scored by either team while the player is on the court is taken into consideration. Net rating has a correlation of 1.0 with wins; a positive season-long net rating indicates a team outscoring its opponent on average. I hoped to determine the best metric predictor of team
regular season win percentage from this regression analysis. The output of the analysis is shown in Table 6 below.

Table 6 - Regression Output from Win Percentage on Net Rating

| Variable | Estimate | T-value | Std. Error | P-value | 95\% Lower | 95\% Upper |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Intercept | 49.98 | 227 | 0.22 | $<\mathbf{2 e - 1 6}$ | 49.54 | 50.41 |
| NetRtg | 2.93 | 68 | 0.04 | $\mathbf{< 2 e - 1 6}$ | 2.84 | 3.01 |
|  |  |  |  |  |  |  |
| R-Squared $=0.94$ |  |  |  |  |  |  |
| Residual standard error $=3.81$ |  |  |  |  |  |  |
| $\mathrm{~N}=300$ |  |  |  |  |  |  |
| Residuals: |  |  |  |  |  |  |
| Min | 1Q | Median | 3Q | Max |  |  |
| -11.91 | -2.74 | 0.01 | 2.69 | 10.57 |  |  |

P-values bolded are less than 0.05

As evident from Table 6, net rating alone is an excellent predictor of team win percentage. The predicted win percentages of $50 \%$ of teams are within approximately 2.7 percentage points above or below their actual regular season win percentages. Though net rating is a strong predictor of team wins, the variable's usefulness in isolating player contributions to team win percentage is limited. As shown by the metric's definition, net rating is affected by the actions of all ten players on the court at any given time. Therefore, it is difficult to attribute the changes in net rating to any one player on the court.

Including only players who had played in ten or more games this season, Los Angeles Clippers guard CJ Wilcox led the NBA in net rating with 29.0 as of February $22^{\text {nd }}, 2016$. Wilcox's high net rating may have been due to the fairly small sample size of his performance (he only averaged about four minutes per game). Second to Wilcox was, predictably, Stephen Curry with 21.1, followed by many San Antonio Spurs, Golden State Warriors, and Oklahoma

City Thunder players. In fact, of all players who had played in ten or more games, only three players in the top 25 in net rating were not on the Spurs, Warriors, or Thunder. These top 25 net rating players included stars from the three highest win percentage teams in the Western Conference, such as Draymond Green, Kawhi Leonard, and Kevin Durant, but also included role players such as Festus Ezeli and Kyle Anderson. Because Ezeli and Anderson often play in lineups with the aforementioned stars, their net ratings were higher than if they had played on different teams; their high net ratings were representative of their teammates, not themselves. LeBron James is much more talented than Ezeli and Anderson, yet his net rating was lower than their ratings. Due to this inability to isolate a player's true talents, I am not including net rating in the model predicting team regular season win percentage.

Player Impact Estimate ("PIE") is an advanced metric that "is an estimate of a player's or team's contributions and impact on a game. PIE shows what $\%$ of game events did that player or team achieve" (NBA.com).

$$
\begin{align*}
\text { PIE }=(\text { PTS } & + \text { FGM }+ \text { FTM }- \text { FGA }- \text { FTA }+ \text { DREB }+(.5 \times \text { OREB })+\text { AST }+ \text { STL } \\
& +(.5 \times \text { BLK })-\mathrm{PF}-\mathrm{TO}) \div(\mathrm{GmPTS}+\mathrm{GmFGM}+\mathrm{GmFTM}-\mathrm{GmFGA} \\
& -\mathrm{GmFTA}+\mathrm{GmDREB}+(.5 \times \mathrm{GmOREB})+\mathrm{GmAST}+\mathrm{GmSTL}  \tag{5}\\
& +(.5 \times \mathrm{GmBLK})-\mathrm{GmPF}-\mathrm{GmTO})
\end{align*}
$$

As evident from Equation 5, PIE records player or team activity in all aspects of the game. Shown below are the results from regressing win percentage on PIE.

$$
\begin{equation*}
\text { Winning } \%=\alpha_{0}+\alpha_{1} \text { PIE } \tag{6}
\end{equation*}
$$

Table 7 - Regression Output from Equation 6

| Variable | Estimate | T-value | Std. Error | P-value | 95\% Lower | 95\% Upper |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Intercept | -165.29 | -43.98 | 3.76 | $<\mathbf{2 e - 1 6}$ | -172.68 | -157.89 |
| PIE | 4.31 | 57.42 | 0.07 | $<\mathbf{2 e - 1 6}$ | 4.16 | 4.45 |

R-Squared $=0.92$
Residual standard error $=4.46$
$\mathrm{N}=300$
Residuals:
Min 1Q Median 3Q Max
$\begin{array}{lllll}-11.07 & -2.96 & -0.04 & 3.00 & 14.45\end{array}$
P -values bolded are less than 0.05

Similar to net rating, PIE is an excellent predictor of team win percentage. The coefficient is positive, sensibly indicating that as a team's PIE increases, or the team becomes more efficient, the team will win more games. The coefficient's magnitude estimates that if the team's PIE increases by one unit, then holding all else constant, the team will win between 3 or 4 more games during the regular season. The r-squared is extremely high at 0.92 , indicating that approximately $92 \%$ of the variation in team regular season win percentage can be explained by only PIE. In total, PIE is a strong predictor for team regular season win percentage, and is used later when calculating appropriate player salaries. However, I do not include PIE in Equation 4, as incorporating PIE into the equation forces each of the other variables to become statistically insignificant. Thus, I use Equation 4 and 6 separately and compare results.

## B. Principal Components Analysis

The second methodology to forming a linear regression equation that predicts team regular season win percentage from multiple per game metrics is called "principal components analysis". Principal components analysis does not regress the response variable directly on the explanatory variables. Instead, win percentage is regressed on the principal components selected within the explanatory variables. Though it is difficult to understand each component from a basketball perspective, by finding the components that make up each of the collected metrics, I effectively limit the number of predictors in my regression model - avoiding the multicollinearity that would otherwise exist.

The results of the principal components analysis are shown in Table 8, where V1 through V9 represent the nine components that explain the most variance of the explanatory variables. In total, these nine components explain approximately $75 \%$ of the variance of the independent variables. After converting the teams' statistical values into the principal components V1 through V9, team win percentage was regressed on all of the teams' principal components.

Table 8 - Principal Components Analysis


P -values bolded are less than 0.05

Though the principal components successfully explain the variance amongst the 76 recorded metrics, not all of these 76 metrics are useful in explaining team win percentage. As evident from Table 8 , only three of the nine components are statistically significant, with p values less than 0.05 . These results indicate the usefulness of "manual elimination", as then the statistically insignificant or unhelpful predictors are ignored when predicting team regular season win percentage.

To test the strength of Equation 4 and 6 in predicting future win percentage, I estimated each team's current (as of March 14 ${ }^{\text {th }}, 2016$ ) 2015-2016 regular season win percentage using actual team statistics. On average, ' $15-$ ' 16 teams had played 65.8 games by March $14^{\text {th }}$. The predicted win percentages for each team using Equation 4 and 6, along with the differences from actual win percentage, are shown in Table 9.

Table 9 - Predicted vs Actual Win Percentages

| Team | Actual W\% | Equation 4 Predicted W\% | Difference | Equation 6 Predicted W\% | Difference |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Atlanta | 56.70\% | 45.22\% | 11.48\% | 61.60\% | -4.90\% |
| Boston | 59.10\% | 36.84\% | 22.26\% | 61.17\% | -2.07\% |
| Brooklyn | 27.30\% | 28.95\% | -1.65\% | 34.48\% | -7.18\% |
| Charlotte | 56.90\% | 42.94\% | 13.96\% | 52.99\% | 3.91\% |
| Chicago | 50.00\% | 43.07\% | 6.93\% | 47.39\% | 2.61\% |
| Cleveland | 72.30\% | 53.61\% | 18.69\% | 63.75\% | 8.55\% |
| Dallas | 50.00\% | 42.38\% | 7.62\% | 52.13\% | -2.13\% |
| Denver | 42.40\% | 33.72\% | 8.68\% | 43.52\% | -1.12\% |
| Detroit | 51.50\% | 33.70\% | 17.80\% | 43.52\% | 7.98\% |
| Golden State | 90.80\% | 72.73\% | 18.07\% | 80.11\% | 10.69\% |
| Houston | 50.00\% | 32.38\% | 17.62\% | 41.37\% | 8.63\% |
| Indiana | 53.00\% | 30.84\% | 22.16\% | 55.57\% | -2.57\% |
| L.A. Clippers | 64.60\% | 51.43\% | 13.17\% | 62.46\% | 2.14\% |
| L.A. Lakers | 20.90\% | 12.75\% | 8.15\% | 18.98\% | 1.92\% |
| Memphis | 59.10\% | 40.05\% | 19.05\% | 50.41\% | 8.69\% |
| Miami | 57.60\% | 45.23\% | 12.37\% | 59.88\% | -2.28\% |
| Milwaukee | 43.30\% | 26.99\% | 16.31\% | 43.95\% | -0.65\% |
| Minnesota | 31.80\% | 31.96\% | -0.16\% | 46.10\% | -14.30\% |
| New Orleans | 36.90\% | 29.75\% | 7.15\% | 39.21\% | -2.31\% |
| New York | 41.20\% | 31.44\% | 9.76\% | 45.24\% | -4.04\% |
| Oklahoma City | 66.70\% | 59.80\% | 6.90\% | 67.63\% | -0.93\% |
| Orlando | 43.10\% | 29.03\% | 14.07\% | 43.09\% | 0.01\% |
| Philadelphia | 13.60\% | 6.46\% | 7.14\% | 18.98\% | -5.38\% |
| Phoenix | 25.80\% | 13.04\% | 12.76\% | 26.73\% | -0.93\% |
| Portland | 52.20\% | 48.68\% | 3.52\% | 47.39\% | 4.81\% |
| Sacramento | 38.50\% | 31.39\% | 7.11\% | 46.53\% | -8.03\% |
| San Antonio | 84.80\% | 68.07\% | 16.73\% | 87.86\% | -3.06\% |
| Toronto | 68.80\% | 56.00\% | 12.80\% | 61.17\% | 7.63\% |
| Utah | 47.00\% | 38.77\% | 8.23\% | 50.41\% | -3.41\% |
| Washington | 46.20\% | 19.66\% | 26.54\% | 43.52\% | 2.68\% |

Differences highlighted in green are within $10 \%$

As can be seen from Table 9, Equation 6 does a much better job of estimating 2015-2016 team win percentages. Only two of the thirty teams' win percentages are not estimated within ten percentage points using Equation 6. Interestingly, Equation 4 estimates the win percentage of one of these two teams, the Minnesota Timberwolves, within 0.16 percentage points - by far the most closely estimated win percentage by Equation 4. On average, Equation 4 overestimates the actual win percentage by 12.17 percentage points, while Equation 6 on average overestimates the actual win percentage by 0.16 percentage points. The fact that Equation 6 is a much better predictor of team regular season win percentage is consistent with my regression analysis results, as the r -squared of Equation 6 is 0.92 with a residual standard error of 4.46 , while the r -squared of Equation 4 is 0.82 with a residual standard error of 6.61 .

As Table 9 shows, Equation 4 overestimates 28 of the 30 teams' regular season win percentages. This may be due to the fact that average per game statistical totals for teams are different now than they were ten years ago. Thus, when using teams' statistics over ten seasons to form the coefficient estimates for predicting team regular season win percentage, the coefficient estimates do not take into account the new ease in acquiring, for instance, a higher $\mathrm{AST} / \mathrm{TO}$ ratio (the $\mathrm{AST} / \mathrm{TO}$ ratio increased from a team average of 1.44 in the 2005-2006 season to an average of 1.55 in '14-'15). Because statistics like AST/TO are higher now than during the ten season span used in the regression analysis, teams are expected to achieve win percentages higher than they actually will during the regular season.

Equation 6, meanwhile, overestimates 13 of the 30 teams' regular season win percentages - a much more symmetric or randomly distributed error. Though teams are totaling higher or lower statistical numbers for specific metrics, all teams are doing so together. Thus, there is no change in efficiency from a statistical viewpoint for two teams during a game. As a result, it is
expected that Equation 6, or really PIE for that matter, would over and underestimate team regular season win percentage approximately the same.

## VII. Player Findings

Based upon my analysis of team performance, I sought to predict each of the explanatory variables in Equations 4 and 6 for players in their next season. I began my investigation predicting the variables in Equation 4.

## A. Equation 4

I started by analyzing the 92 players who played all ten seasons in the dataset. Though viewing only ten-year players greatly reduced my number of observations, I hoped to use the most information possible on a player by extending a player's previous history to 10 seasons. I regressed each player's 2014-2015 metric on his previous season's number of wins, current age, age squared, and the per game averages of the same metric for the previous 9 seasons. I included the player's previous season's number of wins in order to take into account a player's history of strong performance. Though I hoped to use a player's missing games (or 82 minus number of played games) as a variable representing a player's injury history, I decided to leave missing games out of the regression equation. Too many players missed games during the regular season due to coaching decisions, and I would have confused a player's injury woes with poor performance or being a weaker player on a strong roster. Thus, there is no explanatory variable that takes into account a player's injury history in my player prediction equations.

After regressing, for example, the '14-' 15 FTM variable on last regular season's wins, current age, age squared, and the previous nine seasons' FTM, I performed a Durbin-Watson

Test on each regression equation to test for autocorrelation amongst the errors terms. None of the Durbin-Watson Tests posted p-values less than 0.02 , and I proceeded with my analysis that there was no autocorrelation in the error terms. I followed a backwards elimination process of removing variables, eliminating the metric from the earliest season if it were statistically insignificant ( $p$-value $>0.05$ ), forming the regression equation without the variable, and continuing the process of elimination. After reaching a point where all historical metrics were significant, I then removed previous regular season wins if it were insignificant, age squared, and then age. If age and age squared were significant when one was kept in the model, but not when both were in the model, I put each of the variables in the model; when both of the variables were left in the model, the r-squared of the model increased.

The equations for FTM, PF, PACE, REB, and opponent FG\% less than 5 feet include only one year of experience in predicting the next season's metric. The AST/TO variable was best predicted when using the AST/TO per game values from one year and three years ago. Only the opponent field goal percentage from 10-14 feet from two seasons ago was statistically significant in predicting next season's opponent FG\% from 10-14 feet. Interestingly, eFG\% was best predicted by the eFG\% from six seasons prior, while opponent field goal percentage from 20-24 feet was best predicted from per game averages from six and eight seasons ago.

Realistically, though, this is not very helpful when trying to appropriately pay NBA free agents, as NBA players play on average 4.8 years in the NBA (Nelson, 2013). In addition, the contracts of NBA first round draft picks are guaranteed for only two years, while contracts for second round picks typically last for two years as well. Thus, teams often only have two years to analyze players before extending these young free agents a contract offer. For players entering free agency later in their careers, examining the eFG\% or opponent FG\% from 20-24 feet of the
player six years ago is plausible and potentially useful for predicting the eFG\% and opponent FG\% from 20-24 feet of the player for the following season.

Because six of the nine variables were best predicted using data from only the two previous seasons, rookies often play only two seasons in the NBA before entering free agency, and, because I wanted to increase my number of observations, I chose to analyze all players in the NBA who played for three straight seasons. The third season's metrics acted as dependent variables, while the two earlier seasons' metrics were predictors.

Many NBA players missed entire seasons but later returned to the league to play. Such reasons for missing entire seasons included injury and playing in a league abroad. For players who returned after missing whole seasons, I treated their performance history as two separate players before and after the missing season, effectively increasing my number of observations. Because I did not care which player I was analyzing, this data manipulation was doable.

In rare instances, players did not record an opponent field goal percentage from a specific distance. This occurred if the player played extremely limited minutes during the season. In these instances, I gave the player the average opponent field goal percentage from that distance from the 2014-2015 season as his opponent field goal percentage.

The equations for predicting each of the explanatory variables found in Equation 4 are shown below, where I followed the backwards selection process indicated earlier using only two seasons to predict the third season. All of the Durbin-Watson Tests showed no autocorrelation amongst the errors terms, with p-values greater than 0.02 . In each equation, the " .1 " or ". 2 " next to a variable name indicates that the metric is from one or two seasons ago respectively. The "W.1" variable represents the player's number of regular season wins from the previous season.

Table 10 displays the r-squared and standard residual errors for the equations predicting each of the player statistics.

$$
\begin{gathered}
\text { FTM }=\alpha_{0}+\alpha_{1} \text { Age }+\alpha_{2} \text { Age squared }+\alpha_{3} \text { FTM. } 1+\alpha_{4} \text { FTM. } 2 \\
\mathrm{eFG} \%=\alpha_{0}+\alpha_{1} \mathrm{~W} .1+\alpha_{2} \text { Age }+\alpha_{3} \text { Age squared }+\alpha_{4} \mathrm{eFG} \% .1+\alpha_{5} \mathrm{eFG} \% .2 \\
\text { PF }=\alpha_{0}+\alpha_{1} \mathrm{~W} .1+\alpha_{2} \text { Age }+\alpha_{3} \text { Age squared }+\alpha_{4} \text { PF. } 1+\alpha_{5} \text { PF. } 2 \\
\text { PACE }=\alpha_{0}+\alpha_{1} \mathrm{~W} .1+\alpha_{2} \text { Age }+\alpha_{3} \text { Age squared }+\alpha_{4} \text { PACE. } 1 \\
\text { AST } / \mathrm{TO}=\alpha_{0}+\alpha_{1} \text { AST } / \text { TO. } 1+\alpha_{2} \text { AST } / \text { TO. } 2 \\
\text { REB }=\alpha_{0}+\alpha_{1} \mathrm{~W} .1+\alpha_{2} \text { Age }+\alpha_{3} \text { Age squared }+\alpha_{4} \text { REB. } 1+\alpha_{5} \text { REB. } 2 \\
\text { Opp FG } \%<5 \mathrm{ft}=\alpha_{0}+\alpha_{1} \mathrm{~W} .1+\alpha_{2} \text { Opp FG } \%<5 \mathrm{ft} .1
\end{gathered}
$$

Opp 10-14ft FG $\%=\alpha_{0}+\alpha_{1} \mathrm{~W} .1+\alpha_{2}$ Age $+\alpha_{3}$ Age squared $+\alpha_{4}$ Opp 10-14ft FG $\% .1+$ $\alpha_{5}$ Opp 10-14ft FG\%. 2

Opp 20-24ft FG $\%=\alpha_{0}+\alpha_{1} \mathrm{~W} .1+\alpha_{2}$ Age $+\alpha_{3}$ Age squared $+\alpha_{4}$ Opp 20-24ft FG $\% .1+$ $\alpha_{5}$ Opp 20-24ft FG\%. 2

Table 10 - Regression Results from Predicting Player Metrics

| Variable | R-squared | Residual Standard Error |
| :--- | :---: | :---: |
| FTM | 0.74 | 0.62 |
| eFG\% | 0.13 | 10.51 |
| PF | 0.49 | 0.49 |
| PACE | 0.08 | 3.30 |
| AST/TO | 0.35 | 0.95 |
| REB | 0.74 | 1.15 |
| Opp FG\% < 5ft | 0.02 | 7.2 |
| Opp 10-14ft FG\% | 0.01 | 10.57 |
| Opp 20-24ft FG\% | 0.00 | 7.26 |

Using these equations, I then predicted the 2015-2016 regular season statistics of current players who have played for the past two seasons. Table 11 shows the average and median difference of the 2015-2016 actual results minus the 2015-2016 predicted player results.

Table 11 - 2015-2016 Differences in Actual and Predicted Player Metrics

|  | FTM | eFG\% | PF | PACE | AST/TO | REB | Opp | Opp | Opp |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| FG\% $<$ |  |  |  |  |  |  |  |  |  |
| $10-14 \mathrm{ft}$ |  |  |  |  |  |  |  |  |  |
| $20-24 \mathrm{ft}$ |  |  |  |  |  |  |  |  |  |
| 5 ft | FG\% | FG\% |  |  |  |  |  |  |  |
| Avg | -1.74 | 1.89 | 0.11 | 2.25 | 0.00 | 0.01 | -0.07 | 1.33 | -0.83 |
| Median | -1.52 | 1.74 | 0.14 | 2.14 | -0.04 | -0.05 | -0.19 | 1.52 | -0.40 |

The average difference between actual and predicted results is the largest for PACE. This is because PACE is the variable most dependent on team roster and coaching staff. If a player moved to a team with a drastically different coaching philosophy, his PACE would change regardless of his personal change in skill set. Since the model predicting PACE cannot anticipate a change in roster or coach, the difference between actual and predicted results is largest on average for PACE.

The negative average difference for FTM is expected. The average number of FTM per player per game has decreased from 1.61 to 1.39 from the ' $05-$ ' 06 to ' $14-$ ' 15 season. As of March $14^{\text {th }}, 2016$, players averaged 1.43 FTM per game for the ' $15-$ ' 16 regular season. Because of the decline in FTM over the past ten seasons, the model overestimates the number of FTM for players in future seasons. The positive average difference for $\mathrm{eFG} \%$ is expected due to the trend shown in Figure 1. As players make more three pointers per game, their eFG\% rises due to the higher value of three pointers compared to two pointers (see Equation 2). As long as players continue to make more three pointers than they did in the previous ten seasons, the model is expected to underestimate the effective field goal percentage for players. Lastly, despite their fairly low r-squared values and high residual standard errors shown in Table 10, the models for opponent field goal percentages did a strong job of predicting the next season's values.

## B. Equation 6

For comparative purposes, when trying to predict a player's next season's PIE, I also used players who played three straight seasons. I followed the same backwards selection process as indicated above, removing variables if they were statistically insignificant with a p-value greater than 0.05 . In addition, I kept both age and age squared in the equation if both of the
variables were significant when the other was removed from the regression. The equation for predicting PIE is shown below.

$$
\begin{equation*}
\text { PIE }=\alpha_{0}+\alpha_{1} \text { W. } 1+\alpha_{2} \text { Age }+\alpha_{3} \text { Age squared }+\alpha_{4} \text { PIE. } 1+\alpha_{5} \text { PIE. } 2 \tag{7}
\end{equation*}
$$

The Durbin-Watson Test's p-value for Equation 7 was 0.26, indicating that there was not sufficient evidence to suggest that there was autocorrelation amongst the error terms. Table 12 displays the r-squared and residual standard error of Equation 7, while Table 13 displays the average and median difference between the actual and predicted PIE values for players during the 2015-2016 regular season.

Table 12 - Regression Results from Equation 7

|  | R-squared | Residual Standard Error |
| :--- | :---: | :---: |
| PIE | 0.13 | 7.24 |

Table 13 - 2015-2016 Differences in Actual and Predicted Player PIE

| Average | 0.44 |
| :--- | :---: |
| Median | 0.52 |

Though the model predicting PIE had a fairly low r-squared and high residual standard error, as evident from Table 12, the model did a great job predicting a player's next season's PIE.

Because there is no constant trend in player PIE over the past ten seasons, similar to the results found in Table 9, there is no strong over or underestimation for the 2015-2016 season.

## VIII. Assigning Dollar Values

Using the models above, I cannot predict player performance for players who have played one or zero previous seasons in the NBA. Unfortunately, every roster in the NBA for the 2015-2016 regular season had at least three players that fell within this criteria as of March $14^{\text {th }}$, 2016. To solve this issue, I gave second-year players their statistics from the previous season. I gave rookies drafted during the first round the median value of a statistic from player performance during the 2014-2015 season. I gave rookies drafted in the second round, undrafted rookies, or players coming from overseas the tenth percentile value of a statistic from player performance during the 2014-2015 season. When calculating team eFG\%, opponent field goal percentages, AST/TO, and PACE, I needed to assume the minutes per game each player would receive in order to determine the team-weighted averages for these metrics. I assumed constant minutes per game for all players from their last season if they played in the previous season. All first round picks were assumed to have the median number of minutes per game from the 20142015 season; all other first year players were assumed to have the tenth percentile number of minutes per game from the 2014-2015 season.

When calculating PIE using Equation 5 for a team, the team becomes the "player" in the analysis. However, because most players on a team play a unique number of minutes, and, the "game statistics" in Equation 5 consider only those statistics in which a player was on the court, players on the same team do not have the same denominator in Equation 5. Furthermore, the team plays the entire game, and as such does not have the same Equation 5 denominator as any
player. As a result, a team's Player Impact Estimate is not the sum of its players' Player Impact Estimates; I needed to find a relationship between the sum of the players' PIE for a team and the team's own PIE. I summed the players' PIE for all 30 teams as of March $14^{\text {th }}, 2016$ during the '15-' 16 regular season, and divided this total by the team's own PIE on that day. The average multiple was 2.73 , with a minimum multiple of 2.09 and a maximum multiple of 3.56 . I chose 2.73 as the multiple to use when converting player PIE to team PIE going forward.

After predicting or assuming (as indicated above) player performance for every player during the 2015-2016 season, I calculated the per game averages for each of the metrics in Equation 4 and Equation 6 for the 30 NBA teams. I then subtracted every team's actual 20152016 statistics from their predicted ' $15-$ ' 16 statistics, and divided every team's predicted statistics by their actual statistics. As expected, the free throws made per game metric had the greatest difference and multiple between predicted and actual regular season values. The average FTM difference was 18.89 , while the median FTM difference was 19.40 ; the average FTM multiple was 2.06, while the median FTM multiple was 2.04 . Rebounds per game proved to have the second greatest difference, with an average difference of 8.36 and a median difference of 8.39. Personal fouls had the second highest multiple, with an average multiple of 1.30 and a median multiple of 1.31.

I corrected each of the predicted values using their average and median differences and multiples, effectively providing me with five team per game values for each metric. I then calculated the team win percentages using the team per game averages corrected by the average differences, median differences, average multiples, median multiples, and no correction. Column 1 of Table 14 shows the predicted minus the actual win percentages for each of the 30 teams during the '15-' 16 regular season, while columns 2 through 5 show residuals (predicted - actual)
with the predicted statistics corrected by the average differences, median differences, average multiples, and median multiples respectively.

Table 14 - Predicted Team Statistics for ' 15 -' 16 from Equation 4 Minus Actual Statistics

|  | 1 | 2 | 3 | 4 | 5 |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Atlanta | $37.27 \%$ | $-11.85 \%$ | $-12.51 \%$ | $-11.39 \%$ | $-10.10 \%$ |
| Boston | $25.49 \%$ | $-23.63 \%$ | $-24.29 \%$ | $-19.92 \%$ | $-18.72 \%$ |
| Brooklyn | $13.62 \%$ | $-35.49 \%$ | $-36.15 \%$ | $-22.14 \%$ | $-21.06 \%$ |
| Charlotte | $48.97 \%$ | $-0.14 \%$ | $-0.80 \%$ | $-3.80 \%$ | $-2.44 \%$ |
| Chicago | $41.09 \%$ | $-8.02 \%$ | $-8.68 \%$ | $-10.06 \%$ | $-8.80 \%$ |
| Cleveland | $44.12 \%$ | $-4.99 \%$ | $-5.65 \%$ | $-13.03 \%$ | $-11.61 \%$ |
| Dallas | $28.02 \%$ | $-21.10 \%$ | $-21.76 \%$ | $-20.94 \%$ | $-19.58 \%$ |
| Denver | $-6.62 \%$ | $-55.73 \%$ | $-56.39 \%$ | $-43.88 \%$ | $-42.85 \%$ |
| Detroit | $12.33 \%$ | $-36.79 \%$ | $-37.45 \%$ | $-28.40 \%$ | $-27.32 \%$ |
| Golden State | $22.84 \%$ | $-26.28 \%$ | $-26.94 \%$ | $-24.66 \%$ | $-23.37 \%$ |
| Houston | $74.84 \%$ | $25.72 \%$ | $25.06 \%$ | $13.76 \%$ | $15.10 \%$ |
| Indiana | $29.53 \%$ | $-19.58 \%$ | $-20.24 \%$ | $-23.82 \%$ | $-22.48 \%$ |
| L.A. Clippers | $76.56 \%$ | $27.44 \%$ | $26.78 \%$ | $11.29 \%$ | $12.84 \%$ |
| L.A. Lakers | $33.50 \%$ | $-15.62 \%$ | $-16.27 \%$ | $-13.71 \%$ | $-12.52 \%$ |
| Memphis | $34.57 \%$ | $-14.55 \%$ | $-15.21 \%$ | $-15.34 \%$ | $-14.11 \%$ |
| Miami | $34.25 \%$ | $-14.86 \%$ | $-15.52 \%$ | $-18.19 \%$ | $-16.98 \%$ |
| Milwaukee | $39.36 \%$ | $-9.76 \%$ | $-10.42 \%$ | $-5.96 \%$ | $-4.79 \%$ |
| Minnesota | $37.07 \%$ | $-12.04 \%$ | $-12.70 \%$ | $-6.71 \%$ | $-5.58 \%$ |
| New Orleans | $79.35 \%$ | $30.23 \%$ | $29.57 \%$ | $23.26 \%$ | $24.62 \%$ |
| New York | $29.44 \%$ | $-19.68 \%$ | $-20.34 \%$ | $-18.29 \%$ | $-17.07 \%$ |
| Oklahoma City | $66.12 \%$ | $17.00 \%$ | $16.34 \%$ | $2.67 \%$ | $4.02 \%$ |
| Orlando | $33.37 \%$ | $-15.75 \%$ | $-16.41 \%$ | $-12.68 \%$ | $-11.46 \%$ |
| Philadelphia | $-9.35 \%$ | $-58.47 \%$ | $-59.13 \%$ | $-35.13 \%$ | $-34.28 \%$ |
| Phoenix | $47.69 \%$ | $-1.43 \%$ | $-2.09 \%$ | $2.80 \%$ | $3.91 \%$ |
| Portland | $13.82 \%$ | $-35.29 \%$ | $-35.95 \%$ | $-26.92 \%$ | $-25.85 \%$ |
| Sacramento | $56.93 \%$ | $7.81 \%$ | $7.16 \%$ | $2.43 \%$ | $3.72 \%$ |
| San Antonio | $23.41 \%$ | $-25.70 \%$ | $-26.36 \%$ | $-32.91 \%$ | $-31.47 \%$ |
| Toronto | $57.41 \%$ | $8.29 \%$ | $7.63 \%$ | $-0.18 \%$ | $1.18 \%$ |
| Utah | $31.60 \%$ | $-17.51 \%$ | $-18.17 \%$ | $-12.75 \%$ | $-11.57 \%$ |
| Washington | $56.11 \%$ | $6.99 \%$ | $6.33 \%$ | $0.44 \%$ | $1.78 \%$ |
| Average | $37.09 \%$ | $-12.03 \%$ | $-12.69 \%$ | $-12.14 \%$ | $-10.90 \%$ |
| Median | $34.41 \%$ | $-14.71 \%$ | $-15.36 \%$ | $-12.89 \%$ | $-11.59 \%$ |
| Standard Dev | $21.71 \%$ | $21.71 \%$ | $21.71 \%$ | $15.13 \%$ | $15.21 \%$ |
| Cils | $10 \%$ |  |  |  |  |

Cells highlighted in green are within $10 \%$

I conducted the same process for Equation 6 as well. Table 15 shows the results.

Table 15 - Predicted Team Statistics for '15-' 16 from Equation 6 Minus Actual Statistics

|  | 1 | 2 | 3 | 4 | 5 |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Atlanta | $-11.00 \%$ | $13.19 \%$ | $12.02 \%$ | $16.02 \%$ | $13.78 \%$ |
| Boston | $-20.25 \%$ | $3.94 \%$ | $2.77 \%$ | $5.33 \%$ | $3.21 \%$ |
| Brooklyn | $-22.25 \%$ | $1.94 \%$ | $0.78 \%$ | $-0.60 \%$ | $-2.40 \%$ |
| Charlotte | $-15.62 \%$ | $8.57 \%$ | $7.41 \%$ | $10.55 \%$ | $8.38 \%$ |
| Chicago | $-28.68 \%$ | $-4.49 \%$ | $-5.66 \%$ | $-4.81 \%$ | $-6.79 \%$ |
| Cleveland | $-51.66 \%$ | $-27.47 \%$ | $-28.63 \%$ | $-28.22 \%$ | $-30.17 \%$ |
| Dallas | $-20.45 \%$ | $3.74 \%$ | $2.57 \%$ | $4.19 \%$ | $2.15 \%$ |
| Denver | $-38.10 \%$ | $-13.92 \%$ | $-15.08 \%$ | $-16.71 \%$ | $-18.48 \%$ |
| Detroit | $-29.21 \%$ | $-5.02 \%$ | $-6.18 \%$ | $-5.41 \%$ | $-7.39 \%$ |
| Golden State | $-29.22 \%$ | $-5.03 \%$ | $-6.20 \%$ | $-0.54 \%$ | $-2.92 \%$ |
| Houston | $-17.30 \%$ | $6.89 \%$ | $5.72 \%$ | $7.83 \%$ | $5.75 \%$ |
| Indiana | $-27.94 \%$ | $-3.75 \%$ | $-4.92 \%$ | $-3.88 \%$ | $-5.88 \%$ |
| L.A. Clippers | $-21.77 \%$ | $2.41 \%$ | $1.25 \%$ | $4.46 \%$ | $2.29 \%$ |
| L.A. Lakers | $-60.75 \%$ | $-36.57 \%$ | $-37.73 \%$ | $-44.91 \%$ | $-46.22 \%$ |
| Memphis | $-7.66 \%$ | $16.53 \%$ | $15.36 \%$ | $19.59 \%$ | $17.33 \%$ |
| Miami | $-45.49 \%$ | $-21.30 \%$ | $-22.47 \%$ | $-22.93 \%$ | $-24.80 \%$ |
| Milwaukee | $-10.47 \%$ | $13.71 \%$ | $12.55 \%$ | $14.69 \%$ | $12.61 \%$ |
| Minnesota | $-23.24 \%$ | $0.95 \%$ | $-0.22 \%$ | $-1.17 \%$ | $-3.00 \%$ |
| New Orleans | $9.44 \%$ | $33.62 \%$ | $32.46 \%$ | $36.28 \%$ | $34.06 \%$ |
| New York | $-24.03 \%$ | $0.16 \%$ | $-1.01 \%$ | $-1.03 \%$ | $-2.94 \%$ |
| Oklahoma City | $-30.55 \%$ | $-6.37 \%$ | $-7.53 \%$ | $-4.75 \%$ | $-6.89 \%$ |
| Orlando | $-19.85 \%$ | $4.33 \%$ | $3.17 \%$ | $4.01 \%$ | $2.03 \%$ |
| Philadelphia | $-19.86 \%$ | $4.33 \%$ | $3.17 \%$ | $0.25 \%$ | $-1.42 \%$ |
| Phoenix | $-12.13 \%$ | $12.06 \%$ | $10.90 \%$ | $10.80 \%$ | $8.90 \%$ |
| Portland | $-22.81 \%$ | $1.38 \%$ | $0.21 \%$ | $1.80 \%$ | $-0.24 \%$ |
| Sacramento | $-22.67 \%$ | $1.52 \%$ | $0.36 \%$ | $0.28 \%$ | $-1.62 \%$ |
| San Antonio | $-43.85 \%$ | $-19.66 \%$ | $-20.83 \%$ | $-17.59 \%$ | $-19.77 \%$ |
| Toronto | $-37.17 \%$ | $-12.98 \%$ | $-14.15 \%$ | $-12.14 \%$ | $-14.22 \%$ |
| Utah | $-10.73 \%$ | $13.45 \%$ | $12.29 \%$ | $14.95 \%$ | $12.82 \%$ |
| Washington | $-14.54 \%$ | $9.65 \%$ | $8.48 \%$ | $10.85 \%$ | $8.74 \%$ |
| Average | $-24.33 \%$ | $-0.14 \%$ | $-1.30 \%$ | $-0.10 \%$ | $-2.10 \%$ |
| Median | $-22.46 \%$ | $1.73 \%$ | $0.57 \%$ | $0.27 \%$ | $-1.52 \%$ |
| Standard Dev | $14.15 \%$ | $14.15 \%$ | $14.15 \%$ | $15.60 \%$ | $15.47 \%$ |
| Ciso |  |  |  |  |  |

Cells highlighted in green are within $10 \%$

When using a correction method, Equation 6 does a better job of predicting team win percentage than any of Equation 4's methods. Not using a correction method with Equation 6 predicts the win percentage on average much further away from the actual team win percentage than using any of the correction methods on Equation 6. Method 5 for Equation 6, or dividing the raw predicted statistics by 0.896 (the median multiple between raw predicted statistics and actual statistics), has a greater absolute average residual, absolute median residual, and standard deviation than method 3, and is thus ruled out as the best correction method for Equation 6. Method 4 has the smallest absolute average and median difference between predicted and actual statistics, though its standard deviation is larger than that of methods 2 and 3. Cleveland, L.A. Lakers, Miami, and New Orleans all have residuals of at least 20 percentage points. In the case of New Orleans, the overestimation of the team win percentage is largely due to the number of injuries that the team has suffered this season. As of March $30^{\text {th }}$, New Orleans had already lost seven players for the remainder of the season due to injury. Because of these four large residuals in the sample, I chose the correction method with the smallest median residual - method 4. The multiple used for method 4 is 0.887 . Knowing the best method for forming team per game averages from player per game statistics, I then proceeded with altering team rosters and thus their statistics as if free agents joined the team.

I calculated the average salary value of each position (point guard, shooting guard, small forward, power forward, and center) in the NBA for the 2015-2016 season, taking into account only players who received year long contracts; ten-day contracts, D-league contracts, or the contracts of mid-season acquisitions were ignored. I chose to find the average salary value at each position rather than in the NBA as a whole in order to take into account the supply and demand of player types, as well as the differing importance of each position. It is harder to find a
dependable seven-footer than a six-foot four guard, which is evident in centers' and shooting guards' average salaries of $\$ 5,720,130.53$ and $\$ 3,520,657.45$ respectively; point guards, small forwards, and power forwards had average salaries of $\$ 5,113,057.14, \$ 4,534,743.47$, and $\$ 4,429,419.32$ respectively. Point guards are paid higher than shooting guards and forwards due to their difficult task of leading and managing games.

If I wanted to calculate the appropriate salary for a player resigning with a team, I calculated the predicted statistics for the team if the player were to stay on the team for the next season. For example, if the aforementioned Draymond Green stayed on the Golden State Warriors for the 2015-2016 season and the rest of the roster stayed constant, then the Warriors were projected to win $89.36 \%$ of their games using method 4 shown above (the Warriors happen to have won $89.9 \%$ of their games as of March $20^{\text {th }}, 2016$ ). Without Draymond Green, however, I assumed that the average performance of the players at his position on the Warriors would take his place. For the Warriors, Anderson Varejao, Marreese Speights, James Michael McAdoo, and Kevon Looney also play power forward. As a result, I averaged the statistics of each of these players in order to calculate the performance of the player replacing Green. Because PIE is a measure of efficiency, I did not need to take into account the number of minutes played by the replaced or replacing players. The Warriors' regular season win percentage without Green was $84.23 \%, 5.12 \%$ lower or approximately 4.2 fewer games won than with Green. Because the Warriors won $5.12 \%$ more of their games when they had Green instead of their average player, Green deserved to be paid $5.12 \%$ more of the salary cap than the average performing player at his position in the NBA. Since power forwards received on average $\$ 4,429,419.32$ in salary for the 2015-2016 season, Green deserved to receive $\$ 8.0$ million for the 2015-2016 season,
approximately $\$ 6.3$ million less than the $\$ 14.3$ million Green actually received from the Warriors for the '15-' 16 season.

Center Tristan Thompson, who the Cavaliers argued this offseason was comparable to Draymond Green, increased the Cleveland Cavaliers' predicted 2015-2016 win percentage 5.15 percentages points or 4.2 total wins. Thompson, therefore, deserved a $\$ 9.4$ million salary from Cleveland for the ' $15-$ ' 16 season, approximately $\$ 5.9$ million less than the $\$ 14.3$ million Thompson received this season from the Cavaliers. Though Thompson changed the Cavaliers' win percentage by almost the exact same margin as Green, Thompson deserved over one million dollars more because his primary position is center. The model shows that Cleveland was, in fact, fairly correct in comparing Thompson to Green, and that Thompson still deserved to receive significantly less than his actual salary from the Cavaliers.

I followed a similar methodology when calculating the appropriate salary of players joining new teams. I calculated the win percentage of the team without the free agent by simply predicting the team's per game statistics for the following season assuming a constant roster. To determine the win percentage of the team with the free agent, I added the player to the roster and removed the average performance of the player at his position already on the team. For example, when adding Kevin Durant to the Warriors' 2015-2016 roster (a scary thought that might come true in the summer of 2016), I removed the average production of small forwards Harrison Barnes, Andre Iguodala, and Brandon Rush from the Warriors, while adding the statistics of Kevin Durant. According to the model, if Kevin Durant had joined the Warriors before this season, the Warriors would have registered a win percentage of $98.54 \%$, making the team almost unbeatable with the superstar. More realistically, though, for Durant to join the Warriors, Golden State will need to get rid of Harrison Barnes in order to pay Durant the contract he deserves.

Simply replacing Barnes with Durant would have yielded the Warriors a 92.3 win percentage and approximately 2.4 more wins for this 2015-2016 season.

I followed the above methodologies for an additional 40 of last season's NBA free agents. I chose 40 free agents that were diverse in team, position, and salary characteristics. Table 16 reports the actual 2015-2016 salaries of the players, as well as my calculated appropriate salaries for these players based upon the team with which they signed.

Table 16 - Actual vs Appropriate 2015-2016 Salaries

| Player | Team | Actual Salary <br> $(\mathrm{mil})$ | Appropriate Salary <br> $(\mathrm{mil})$ | Difference <br> $(\mathrm{mil})$ |
| :--- | :---: | :---: | :---: | :---: |
| Quincy Acy | SAC | $\$ 0.98$ | $\$ 4.43$ | $-\$ 3.45$ |
| Al-Farouq <br> Aminu | POR | $\$ 7.50$ | $\$ 10.74$ | $-\$ 3.24$ |
| Alan Anderson | WAS | $\$ 4.00$ | $\$ 2.48$ | $\$ 1.52$ |
| Will Barton | DEN | $\$ 3.33$ | $\$ 7.04$ | $-\$ 3.71$ |
| Aron Baynes | DET | $\$ 6.50$ | $\$ 5.59$ | $\$ 0.91$ |
| Matt Bonner | SAS | $\$ 1.50$ | $\$ 4.14$ | $-\$ 2.64$ |
| Jimmy Butler | CHI | $\$ 15.26$ | $\$ 8.22$ | $\$ 7.04$ |
| Omri Casspi | SAS | $\$ 3.00$ | $\$ 5.16$ | $-\$ 2.16$ |
| Tyson Chandler | PHX | $\$ 13.00$ | $\$ 9.20$ | $\$ 3.80$ |
| Norris Cole | NOP | $\$ 3.04$ | $\$ 6.20$ | $-\$ 3.16$ |
| Jae Crowder | BOS | $\$ 6.00$ | $\$ 7.56$ | $-\$ 1.56$ |
| Toney Douglas | NOP | $\$ 1.16$ | $\$ 0.41$ | $\$ 0.76$ |
| Goran Dragic | MIA | $\$ 14.78$ | $\$ 9.67$ | $\$ 5.11$ |
| Wayne Ellington | BKN | $\$ 1.50$ | $\$ 4.29$ | $-\$ 2.79$ |
| Monta Ellis | IND | $\$ 10.30$ | $\$ 6.07$ | $\$ 4.23$ |
| Kevin Garnett | MIN | $\$ 8.50$ | $\$ 6.86$ | $\$ 1.64$ |
| Marc Gasol | MEM | $\$ 19.70$ | $\$ 10.31$ | $\$ 9.39$ |
| Tyler <br> Hansbrough | CHA | $\$ 1.19$ | $\$ 3.30$ | $-\$ 2.12$ |
| Joe Ingles | UTA | $\$ 2.25$ | $\$ 2.73$ | $-\$ 0.48$ |
| Richard <br> Jefferson | CLE | $\$ 1.50$ | $\$ 1.08$ | $\$ 0.42$ |
| John Jenkins | DAL | $\$ 0.98$ | $\$ 3.08$ | $-\$ 2.10$ |
| Jonas Jerebko | BOS | $\$ 5.00$ | $\$ 5.66$ | $-\$ 0.66$ |
| Cory Joseph | TOR | $\$ 7.00$ | $\$ 8.11$ | $-\$ 1.11$ |


| Shane Larkin | BKN | $\$ 1.50$ | $\$ 5.59$ | $-\$ 4.09$ |
| :--- | :---: | :---: | :---: | :---: |
| Jeremy Lin | CHA | $\$ 2.10$ | $\$ 6.34$ | $-\$ 4.24$ |
| Robin Lopez | NYK | $\$ 13.00$ | $\$ 5.52$ | $\$ 7.48$ |
| Kevin Love | CLE | $\$ 19.50$ | $\$ 11.81$ | $\$ 7.69$ |
| KJ McDaniels | HOU | $\$ 3.33$ | $-\$ 0.60$ | $\$ 3.93$ |
| Khris Middleton | MIL | $\$ 15.00$ | $\$ 5.77$ | $\$ 9.23$ |
| Paul Millsap | ATL | $\$ 19.00$ | $\$ 5.36$ | $\$ 13.64$ |
| Jameer Nelson | DEN | $\$ 4.50$ | $\$ 1.88$ | $\$ 2.62$ |
| Pablo Prigioni | LAC | $\$ 0.98$ | $-\$ 1.41$ | $\$ 2.39$ |
| Austin Rivers | LAC | $\$ 3.10$ | $\$ 5.31$ | $-\$ 2.21$ |
| Glenn Robinson | IND | $\$ 0.85$ | $-\$ 0.43$ | $\$ 1.28$ |
| Luis Scola | IND | $\$ 3.00$ | $\$ 4.55$ | $-\$ 1.55$ |
| Kyle Singler | OKC | $\$ 5.00$ | $\$ 3.22$ | $\$ 1.78$ |
| Jason Terry | HOU | $\$ 1.50$ | $\$ 0.15$ | $\$ 1.35$ |
| Charlie <br> Villanueva | DAL | $\$ 1.50$ | $\$ 3.39$ | $-\$ 1.89$ |
| CJ Watson | ORL | $\$ 5.00$ | $\$ 4.07$ | $\$ 0.93$ |
| Lou Williams | LAL | $\$ 7.00$ | $\$ 6.26$ | $\$ 0.74$ |

The salary differences highlighted in green are those in which the actual salary is greater than the appropriate salary, and vice versa for those highlighted in red. The average actual salary of those players whose appropriate salary is greater than their actual salary is $\$ 3.03$ million, with a median of \$2.63 million. The average actual salary of those players whose appropriate salary is less than their actual salary is $\$ 8.61$ million, with a median of $\$ 6.75$ million. It appears that end-of-bench players are thus underpaid for their contributions to team win percentage, while starters or high-end bench players are overpaid for their contributions to team win percentage. Of those free agents listed in Table 16, the highest appropriate salary for a player is $\$ 11.84$ million for power forward Kevin Love of the Cleveland Cavaliers, while the lowest appropriate salary is negative $\$ 1.41$ million for point guard Pablo Prigioni of the Los Angeles Clippers. Prigioni remains on the Clippers' roster more for his leadership and experience (he is 38 years old) than
for his on-court contributions (he averages 12.6 minutes per game in the 51 out of 73 games he has played this season). Playing point guards Chris Paul and Austin Rivers on the roster instead of Pablo Prigioni appears to be the better move for the Clippers going forward. The other players with negative appropriate salaries, KJ McDaniels and Glenn Robinson, face a similar situation as Prigioni in having strong talent at their respective positions on their roster. McDaniels, a shooting guard, plays behind James Harden, Corey Brewer, Trevor Ariza, and Jason Terry on the Houston Rockets. Meanwhile, Robinson's departure from the Indiana Pacers would pave the way for more minutes from Paul George, Monta Ellis, CJ Miles, and Solomon Hill.

The player with the largest positive difference between actual and appropriate salary is Paul Millsap of the Atlanta Hawks; Millsap has a difference of $\$ 13.64$ million, while his actual salary is $\$ 19$ million. Millsap starts at power forward for the Hawks, who also have Kris Humphries, Mike Muscala, and Mike Scott on the roster at power forward. Humphries, Muscala, and Scott have predicted PIE of 9.6, 10.9, and 10.7 respectively, compared to Millsap's 11.1. Thus, replacing Millsap with the average performance of these forwards would only decrease the Hawks’ win percentage by 1.32 percentage points. Memphis Grizzlies center Marc Gasol has the second largest positive difference between actual and appropriate salary with a difference of $\$ 9.39$ million. Having Gasol increases the Grizzlies' number of wins by approximately 5.4, a fairly large margin during the regular season. Thus, the large positive difference between actual and appropriate salary is really a reflection of Gasol's high actual salary of $\$ 19.70$ million for the 2015-2016 season.

It is interesting to note those players that have an appropriate salary greater than their actual salary. These are players potentially underrated and definitely underpaid by their teams. Point guard Shane Larkin of the Brooklyn Nets is the player with the largest negative difference
between actual and appropriate salaries for the '15-' 16 season. Larkin shares the point guard duties for the Nets with the not very impressive group of Donald Sloan, Jarrett Jack, and Markel Brown. Larkin, though, is not much of a better alternative from the average performance of these point guards; Larkin improves the Nets' win percentage by only 0.67 percentage points. In Larkin's case, his appropriate salary is drawn more from the average salary of point guards than from his performance above the average. On a team with better point guards, for example the Los Angeles Clippers, Larkin would deserve a much lower salary.

Unlike many of the other undervalued players, Al-Farouq Aminu, Jae Crowder, and Cory Joseph are three guys that play big minutes and are still paid actual salaries less than their appropriate salaries. The Portland Trailblazers have a steal in small forward Al-Farouq Aminu. Without Aminu, the Trailblazers would be left to play Allen Crabbe, Maurice Harkless, and Luis Montero at their small forward spot. Having Aminu increases the team's regular season win percentage by 8.86 percentage points. As a result, Aminu should be paid $\$ 10.74$ million for his contributions to win percentage for the Portland Trailblazers. Aminu, though, received a $\$ 7.50$ million salary for the 2015-2016 season. Most likely, Aminu received a lower actual salary because his performance the season before on the Dallas Mavericks was limited from sharing small forward duties with Chandler Parsons, Jae Crowder, and Richard Jefferson.

Except for Jae Crowder, the Boston Celtics have small forwards that are fairly young, inexperienced, and not efficient on the court. Crowder received $\$ 6.00$ million from the Celtics for this season, but because the Celtics would win about 3.5 fewer games without him, Crowder deserves a $\$ 7.56$ million salary for the ' $15-$ ' 16 season. After his trade to the Celtics mid-season last year, Crowder's minutes per game increased. Thus, his time-share in Dallas does not appear to be the root of his lower actual salary.

Meanwhile Cory Joseph, back-up point guard for the Toronto Raptors, has a predicted PIE higher than starting Raptors point guard Kyle Lowry and other bench point guard Delon Wright. If Joseph were to leave the team, the Raptors would lose about 3.5 more games if Lowry and Wright were left to replace Joseph. Joseph's high predicted PIE lends credibility to the idea that he should play more of Kyle Lowry's minutes at point guard.

## IX. Conclusion

When trying to predict team regular season win percentage from actual player statistical performance, the advanced metric PIE is a better predictor of team success than the combination of FTM, eFG\%, REB, AST/TO, PACE, PF, and opponent field goal percentages within five feet, 10-14 feet, and 20-24 feet away from the hoop. Also, when trying to use predicted team values for the above metrics, the nine variables do a worse job of predicting team regular season win percentage than team PIE. The average residual (actual - predicted) for predicting team win percentage from the nine variables in Equation 4 is $12.17 \%$ and $37.09 \%$ using actual statistics and predicted team statistics respectively; the average residual (actual - predicted) for predicting team win percentage from Equation 6 is $0.16 \%$ and $24.33 \%$ using actual and predicted team statistics respectively. The drastic increase in the average residuals from using predicted variables is due to the difficulty in predicting player performance from only the previous two seasons. Player performance can stray from a usual track due to injury, roster changes, coaching changes, or in a positive light, practice. As a result, accurately predicting the performance of all players is a struggle.

These findings indicate that teams win not from necessarily tallying huge statistical totals, but from being more efficient than the opposition. As mentioned previously, teams can
win in a variety of ways - playing quick or slow - and the objective for the team is to find its strategy and be efficient at executing it. I hoped through the combination of PACE and the other metrics in Equation 4 that I could manufacture a team's efficiency of play in comparison to that of the opponent. However, by the very nature of the formula, PIE did a more accurate job of illustrating a team's efficiency in comparison to the opponent's efficiency.

My model estimates players to contribute much fewer wins than the Win Shares model indicates. Even as of March $20^{\text {th }}, 2016$, the players with the top five Win Shares in the unfinished 2015-2016 regular season had greater Win Shares than my model's predicted contributed wins for the entire regular season. Table 17 shows the predicted wins contributed for the entire regular season and the actual Win Shares for players as of March $20^{\text {th }}$.

Table 17 - Actual vs Projected Contributed Wins

|  | Win Shares as of <br> $3 / 20 / 16$ | Projected '15-'16 Contributed <br> Wins |
| :--- | :---: | :---: |
| Stephen Curry | 15.2 | 10.4 |
| Kevin Durant | 12.7 | 11.5 |
| Russell |  |  |
| Westbrook | 12.4 | 8.5 |
| Kawhi Leonard | 12.4 | 7.7 |
| Kyle Lowry | 11.5 | 0.3 |

The projected wins contributed are significantly less than the actual Win Shares of players because the projected wins contributed takes into account the players that will replace the analyzed player, while Win Shares does not. When calculating a player's Win Shares, the team is completely ignored except for its PACE and points per possession - only the statistics of the analyzed player are used. When I project the number of wins contributed by a player, I am
calculating the difference in the number of wins that a team achieves with and without the player, having other players replace the player's performance. In other words, I care about how the team will perform when somebody else plays the minutes that the player would play if he were to not be on the team. Because of the difference in what I am trying to calculate, I expect the projected contributed wins to be less than the number of Win Shares.

Looking at Table 17, it is interesting to note which players are "replaceable" by their team. Kevin Durant's Win Shares are only 1.2 wins greater than his season-long projected contributed wins. These metrics are close in value due to the limited skill of the remaining small forwards on the Oklahoma City Thunder roster if Durant were to leave the team. Kyle Lowry, meanwhile, posts a Win Shares over 38 times higher than his projected contributed wins. As I mentioned previously, the Toronto Raptors, Lowry's team, have a great back-up point guard in Cory Joseph who could replace Lowry if he were to leave. The Raptors also feature recent firstround pick Delon Wright - a player capable of filling in at point guard if Lowry were to leave. Thus, the loss of Lowry does not appear to be too significant for the Raptors.

According to my analysis of paying free agents according to how much they affect a team's regular season win percentage, teams overpay free agent superstars and underpay the last men on the bench. This provides evidence to the existence of the tournament theory in the NBA that Berri and Jewell alluded to in "Wage Inequality and Firm Performance: Professional Basketball's Natural Experiment" (Berri \& Jewell, 2004). Though the NBA allows for mobility of players across teams, there appears to be too great of a supply of lower tier players for teams to increase the salary of bench players. Bench players accept these lower wages with the motivation that they will become stars in the NBA, rather than play abroad under a potentially
higher salary. Meanwhile, there is a limited supply of superstars in the NBA, and these players receive the riches resulting from multiple teams bidding for them.

Despite the fact that my model demonstrates that these superstars are overpaid for their contributions to team win percentage, teams would be wrong to pay superstars the salary that the model dictates so long as bench players are willing to receive salaries lower than their worth. A change in salary demanded by end-of-bench players seems unlikely, and the only such reason why bench players would receive higher salaries is if the minimum league salary were increased by the NBA, but no such move appears imminent.

Teams should focus less on which superstars are deemed overpaid (as all superstars are), and focus on which role players are underpaid for their contributions to team win percentage. These role players may be underpaid because their teams have bad players at their respective positions, or because teams overlook the efficiency of these players in their limited minutes. If these players are underpaid because of the team's poor roster composition, then the team should look to improve its roster at the player's respective position; superstars or good starting players would not be underpaid, and thus the team has room to improve this position. In terms of efficiency, teams should start relying heavily (if they do not already) on a player's PIE when determining his worth to the team, as this study proves that a team's PIE is a great predictor for team regular season win percentage. Front offices can look for players that play at a slow speed and thus do not put up big numbers, but have efficient statistics in limited possessions. Toronto Raptors center Jonas Valanciunas is such a player. The Raptors play at the third slowest PACE in the NBA during the 2015-2016 season, restricting the numbers that Valanciunas can obtain. Though he has a PIE of 14.0 as of March $20^{\text {th }}, 2016$, the $27^{\text {th }}$ highest PIE in the NBA of players averaging 25 or more minutes per game, Valanciunas is only $18^{\text {th }}$ in rebounds per game, $37^{\text {th }}$ in
blocks per game, and $93^{\text {rd }}$ in points per game of all players. The Raptors identified Valanciunas' strong ability to impact the game in a limited number of possessions, and have already reached a hefty extension with the center; the Raptors offered Valanciunas $\$ 14.4$ million for the '16-' 17 season.

All in all, this model provides teams with an alternative and easy-to-use method of valuing free agents. As this model's only inputs are player characteristics and statistics recorded on NBA.com, every team can predict player performance, calculate changes in team performance from roster moves, and determine the appropriate salaries of free agents based upon their estimated contribution to win percentage. If teams were to adopt this model, I would not expect teams to pay superstars less or end-of-bench players more. Rather, I would expect teams to devote more attention in free agency to signing efficient role players - those players that deserve more money than what their statistical totals would dictate.

## References

Basketball-Reference.com. (2016). Sports Reference LLC. Retrieved from http://www.basketball-reference.com

Berri, D. J., \& Jewell, R. T. (2004). Wage Inequality and Firm Performance: Professional Basketball's Natural Experiment. Atlantic Economic Journal, 32(2), 130-39.

Coon, Larry. (2015). NBA Salary Cap/Collective Bargaining Agreement FAQ. Retrieved from http://www.cbafaq.com

ESPN. (2016). ESPN Internet Ventures. Retrieved from http://www.espn.go.com
Franks, A., Miller, A., Bornn, L., \& Goldsberry, K. (2015). Counterpoints: Advanced Defensive Metrics for NBA Basketball. MIT Sloan Sports Analytics Conference. Retrieved from http://www.sloansportsconference.com/

Goldsberry, K., \& Weiss, E. (2013). The Dwight Effect: A New Ensemble of Interior Defense Analytics for the NBA. MIT Sloan Sports Analytics Conference. Retrieved from http://www.sloansportsconference.com/

Harrel, Patrick (2014). "Age in the NBA: Do Older Teams "find Ways to Win Games"?" Princeton Sports Analytics. Retrieved from http://princetonsportsanalytics.com/2014/11/03/age-in-the-nba-do-older-teams-find-ways-to-win-games/

Laws, W. (2015). Predicting the NBA standings for the 2015-2016 season using Win Shares. Sports Illustrated. Retrieved from http://www.si.com/nba/2015/07/16/nba-standings-predictions-hornets-thunder-celtics-blazers-mavericks

Lewis, M. (2004). Moneyball: The Art of Winning an Unfair Game. New York, NY and London, UK: W.W. Norton \& Company, Inc.

Lowe, Z. (2013). Lights, Cameras, Revolution. Grantland. Retrieved from http://grantland.com/features/the-toronto-raptors-sportvu-cameras-nba-analyticalrevolution/

NBA.com. (2016). Turner Sports \& Entertainment Digital Network. Retrieved from http://www.nba.com

Nelson, Jeff. (2013). The Longest Professional Sports Careers. The Roosevelts. Retrieved from http://www.rsvlts.com/2013/07/22/longest-sports-careers/
"Pythagorean Theorem of Baseball." (2015). Baseball-Reference.com. MediaWiki. Retrieved from http://www.baseball-reference.com/bullpen/Pythagorean_Theorem_of_Baseball

Strauss, Ethan S. (2013). Houston Rockets are Officially NBA's 'Moneyball' Team. Bleacher Report. Retrieved from http://bleacherreport.com/articles/1485927-houston-rockets-are-officially-nbas-moneyball-team
"What Is Pythagorean \& Its Sports Betting Impact." Maddux Sports. Retrieved from http://www.madduxsports.com/library/handicapping/what-is-pythagorean-its-sports-betting-impact.html

Glossary (terms and abbreviations taken from NBA.com)

```
%FGA 2pts = Percent of Field Goals Attempted 2-points
%FGA 3pts = Percent of Field Goals Attempted 3-points
%pts 2pt-MR = Percent of Points 2-point Field Goals - Mid Range
%pts 2pts = Percent of Points 2-points
%pts 3pts = Percent of Points 3-points
%pts Fast Break Points = Percent of Points Fast Break Points
%pts FT = Percent of Points Free Throws
%pts Off TO = Percent of Points off Turnovers
%pts PITP = Percent of Points Points in the Paint
+/- = Plus/Minus
10-14ft = 10-14ft from the hoop
15-19ft = 15-19ft from the hoop
20-24ft = 20-24ft from the hoop
25-29ft = 25-29ft from the hoop
2FGM %AST = Percent of 2pt Field Goals Made Assisted
2FGM %UAST = Percent of 2pt Field Goals Made Unassisted
2nd PTS = 2nd Chance Points
3FGM %AST = Percent of 3pt Field Goals Made Assisted
3FGM %UAST = Percent of 3pt Field Goals Made Unassisted
3P% = Three Point Percentage
3PA = Three Pointers Attempted
3PM = Three Pointers Made
5-9ft = 5-9ft from the hoop
AST = Assists
AST Ratio = Assist Ratio
AST/TO = Assist to Turnover Ratio
AST% = Assist %
BLK = Blocks
BLKA = Blocks Against
DefRtg = Defensive Rating
DREB = Defensive Rebounds
DREB% = Defensive Rebound %
eFG% = Effective Field Goal Percentage
FBPs = Fast Break Points
FG% = Field Goal Percentage
FGA = Field Goals Attempted
FGM %AST = Percent of Field Goals Made Assisted
FGM %UAST = Percent of Field Goals Made Unassisted
FGM = Field Goals Made
FT% = Free Throw Percentage
FTA = Free Throws Attempted
FTM = Free Throws Made
Less than 5ft = Less than 5ft from the hoop
NetRtg = Net Rating
```

OffRtg = Offensive Rating
Opp = Opponent
OREB = Offensive Rebounds
OREB\% = Offensive Rebound \%
PACE $=$ Pace
PF = Personal Fouls
PFD = Personal Fouls Drawn
PIE $=$ Player Impact Estimate
PITP $=$ Points in the Paint
PTS $=$ Points
PTS Off TO = Points off Turnover
REB \% = Rebound \%
REB $=$ Total Rebounds
STL $=$ Steals
TO = Turnovers
TO Ratio $=$ Turnover Ratio
TS\% = True Shooting \%

