2014

In defense of defense: a statistical look at roster construction, coaching strategy, and team defense in the National Basketball Association

Luke Ronald Peterson
University of Northern Iowa

IN DEFENSE OF DEFENSE:

A STATISTICAL LOOK AT ROSTER CONSTRUCTION, COACHING STRATEGY, AND TEAM DEFENSE

IN THE NATIONAL BASKETBALL ASSOCIATION

A Thesis Submitted

in Partial Fulfillment

of the Requirements for the Designation

University Honors with Distinction

Luke Ronald Peterson

University of Northern Iowa

December 2014

Introduction

Throughout history, people in all cultures have played games, danced, and engaged in physical activity; but it was not until the industrial revolution and urbanization movement of the 1850s that the modern sport movement began. As Gems, Borish, and Pfister (2008) explained, modern sport is different from traditional cultural sport in three fundamental aspects: the equality of opportunity, the focus on performance and competition, and the setting and keeping of records (Gems, Borish, Pfister, 2008). In the last decade or two, a statistical revolution has swept the playing field and added another fundamental aspect to modern sport – advanced analytics and prediction.

The three major organized sports featured in the United States – baseball, basketball, and football – cover the spectrum of analytic probability. Baseball, with its basic two parts, pitcher and batter, lies at the easy or low end of the analytic spectrum and so it was the first sport to be studied. From early baseball cards first appearing in the 1890s to *The Bill James Historical Baseball Abstracts* of the 1980s to the SABRmetrics of today, baseball coaches and fans everywhere are adding a page on analytic probability to their playbooks. Football lies at the high or difficult end of the spectrum. With its 22 different parts or players involved in every play, football remains largely unexplored and largely out of reach from a statistical perspective. Basketball, however, with its 10 parts or players lies in that sweet spot or Goldilocks zone, not too hot and not too cold. Whether one is a casual fan or astute student of the game, advanced analytics have dramatically changed the way the fans perceive these sports. In other words, the shoebox crammed with baseball cards of my dad's generation has morphed into the computer-based analyzes of my generation.

Despite the many reams of statistics that have been added to the encyclopedia of sports knowledge, there are still many unknowns and many chapters that are incomplete or missing entirely. One of these incomplete chapters is team defense in basketball. While there are a few team axioms that exist, such as rim-protecting centers are valuable and slow point guards are not, the relationship between team characteristics and team defense remains vague, at best. Like the dark side of the moon, the unknown aspects of team defense are seldom seen and remain largely unexplored. My goal in writing this paper is to take a closer look at team defense in the National Basketball Association (NBA). By quantitatively looking at several factors, I hope to gain a better understanding of how the measurable quantities of team make-up affect the abstract qualities of team building. While I may not be the first to attempt the trek nor be the one to voyage the farthest towards this final frontier, I hope to boldly go into the great unknown. By researching and analyzing certain NBA team characteristics, in terms of both roster construction and coaching, I hope to be able to ascertain and perhaps even predict the team's defensive ability.

Problem Purpose

The problem as identified is a lack of knowledge regarding the importance of certain characteristics in the development of a team's defense in basketball. This lack of knowledge affects basketball coaches and/or general managers at every level of the sport. My purpose in identifying and connecting specific characteristics within individual five-man lineups to their defensive efficiency is to provide executives, coaches, and fans with a clearer picture and understanding of how given identified variables affect team defense.

In other words, the big picture question may be asked as follows: How do certain team characteristics, in terms of both roster construction and coaching strategy, affect a team's defensive ability?

Hypotheses to be Tested

First, I identified thirteen quantifiable variables related to team building that would appear to have an effect on team defense in the NBA. These thirteen variables can be generally categorized as either roster construction or coaching strategy with some gray areas in differentiation. These thirteen variables can also be described as priors, or variables that can be determined prior to the playing of any basketball games. While not always possible to use a metric that truly encapsulates the prior characteristics of a team, every means to acquire a metric that did so was exhausted. Roster construction variables that were investigated include: age, height, wingspan, athleticism, experience acquired prior to the NBA, total number of college players, coach continuity, player continuity, and individual Player Efficiency Rating (PER). Coaching strategy variables that were investigated include: pace, minutes played, fouls committed, and team offensive efficiency**.** Since this will be a statistical study, thirteen of the hypotheses to be tested are that each individual variable has no relationship to a lineup's defensive efficiency. An additional hypothesis to be tested is that the variables connected to roster construction are more strongly tied to a lineup's defensive efficiency than the variables connected to coaching strategy.

Literature Review

Just as there are two sides to the moon – the light and the dark side – so there are two sides to basketball – offense and defense. The light side of the moon would represent offense

– easy to see, widely studied, in other words the bright, shining, glamorous side of basketball. After all, the team with the most points at the ending buzzer wins the game. Defense, on the other hand, would be found residing on the dark side of the moon – seldom seen, usually ignored, in other words the dark, workmanlike, drudgery of basketball. For example, sports headlines and photos usually feature offensive moves and outcomes. These light and dark sides are clearly presented in NBA box scores that typically list each player, their field goals made and attempted, free throws made and attempted, and total points scored. Similarly, the team stats listed might include 3-point field goals, players who fouled out, total fouls, and rebounds.

In addition to the myriad offensive box score stats, the toolbox of additional offensive stats has continually gotten larger. One such addition was made by Kirk Goldsberry in 2012 with the introduction of CourtVision. CourtVision provides offensive metrics that quantify the spacing that is inherent to a successful offense. These metrics determine the extent to which a player spreads the floor through the percentage of locations shot from and the efficiency exhibited at those locations (Goldsberry, 2012). Another basketball skill that has been better understood through analytic research is rebounding. A rebound in a box score can only describe one event, a player grabbing a ball after either team missed a shot. The actual skill of rebounding requires much more nuance than a lone tally mark in a stat sheet. Mahaswaran, Chong, Su, and Kwok found that there are actually three dimensions of rebounding and created metrics to describe a player's capability in each facet of rebounding (Mahaswaran, Chong, Su, & Kwok, 2014). These offensive metrics describe the action that takes place on the basketball court, but these offensive metrics have focused on outcomes rather than processes. Perhaps the crown jewel of offensive statistics would be able to describe a player's decision-making. This crown jewel may have been found by Cervone, D'Amour, Bornn, and Goldsberry in their 2014 paper "PointWise." Described as "microeconomics for the NBA," Expected Possession Value evaluates every player movement and decision in terms of the expected point value for that possession (Cervone, D'Amour, Bornn, & Goldsberry, 2014).

In comparison to offensive statistics, defensive analyses and studies have not received much press in the written record of published academic research. One area of team defense that has been researched and studied to some degree, although nowhere close to the attention individual offensive stats have received, is individual defensive stats. For example, the groundbreaking paper of the Sloan Sports Analytics Conference of 2013, "The Dwight Effect," written by Kirk Goldsberry and Eric Weiss, relied on 3D-camera data to address the influence of a rim-protecting player upon opponents' field goal percentages within five feet of the basket (Goldsberry & Weiss, 2013). Another example of a recently conceived metric for individual defense was created by Franks, Miller, Bornn, and Goldsberry and determines the frequency and efficiency with which opposing players shoot in different areas of the court when defended by a particular player. Thus, the spatial component of basketball is taken into account when determining the defensive value of individual players (Franks, Miller, Bornn, & Goldsberry, 2014). So while these papers focused on the defensive side of the game, the research and findings centered around the value of an individual player rather than the value of an archetypal player.

While understanding the value of an individual player is significant, perhaps the more important value is that of a combination of players. After all, basketball is not merely a collection of individuals but two teams squaring off against each other. One study that looked at a collection of players, albeit in the form of two and three man combinations, was conducted by Ayer in 2012. Ayer grouped players into types and examined the effect that different player-type combinations have had on team win totals throughout history. Ayer was also able to address the effect of individual coaches based on the extent to which these coaches exceeded expectations (Ayer, 2012). Beyond specific player combinations, two published papers have used statistical analysis to determine what qualities are common in winning teams. Sampaio, Ibáñez, Lorenzo, and Gomez examined results from the Portuguese Professional League to determine which game-related statistics differ between starters and non-starters in two specific game outcomes: when the better team wins the game and when the worse team wins the game (Sampaio, Ibáñez, Lorenzo, & Gomez, 2006). Ergül, Yavuz, and Yavuz examined NBA teams in order to determine which game-related statistics indicate that a team will make the playoffs (Ergül, Yavuz, & Yavuz, 2014). These studies do well to explain what on-court factors indicate success. However, each study examines only the on-court factors and ignores off-court or prior attributes. Therefore, predicting team success with the findings of these studies is difficult. Ideally, a team could be categorized solely on the make-up of the team rather than waiting to see how the team performed on the court.

Dean Oliver, a pioneer in basketball analysis and recent hire in the Sacramento Kings organization, was one of the first researchers to have published his findings in basketball statistics, specifically findings regarding team characteristics and performance. He has authored *Basketball on Paper: Rules and Tools for Performance Analysis*, published in 2003, as well as *The Journal of Basketball Studies*, an online compilation of basketball research. Oliver looked at the top 25 defensive teams and the worst 25 defensive teams in the NBA from 1974-2002. Comparing these teams to the average team, he found that the best defensive teams have players that are slightly above average in height while averaging half a possession less per game or, in other words, playing at a slightly slower than average pace. He also found that the worst defensive teams had newly assembled coaching staffs and player rosters (Oliver, 2004).

So while there has been some research dealing with an individual's contribution on the defensive side, on-court statistics relating to team success, and categorizing the best and worst team defenses, there remains a sizeable gap in the literature relating to off-court team characteristics in the form of roster construction and coaching strategy. Of course, it is highly likely that more research on this topic has been conducted by NBA organizations but remains unpublished due to teams attempting to achieve a competitive advantage.

Methodology

In order to gain a better understanding of the relationship between the thirteen identified variables and team defense, regression analyzes were utilized. The dependent variable in each analysis was the defensive efficiency of a five-man lineup. In the search for meaningful information, five-man lineups provide more insight into who is actually on the court and the level of success they are achieving than analyzing the entire team. The lineup data will cover six seasons from 2007-08 through 2013-14 as recorded on the official NBA website (www.nba.com/stats). These data will include the defensive efficiency as well as other relevant information for every lineup that has played 400 or more minutes together during each season. While 400 minutes is an arbitrary cut-off, it provides a large enough sample size for each lineup, the equivalent of roughly 8 full games, while still including enough lineups to provide a meaningful data set. I also excluded all lineups from the lockout-shortened 2011-12 season as extenuating circumstances compromised much of that season; therefore, these data are an outlier in regard to the history of the league.

Definitions

In basketball, a possession is defined as a time period where one team has control of the ball until it loses control of the ball to the other team by shooting, turning the ball over, or shooting free throws. Possessions become the name of the game when attempting to understand what is happening on the court because they are the shortest time frame during which meaningful action occurs. Ultimately, the game of basketball boils down to how effective a team is when either they or their opponent has control of the ball. A team's total possessions in any given game are often estimated from end-of-game totals, and in this study the estimated total possession numbers as published by the NBA were used.

Of the variables used in this study, the following four characteristics are self-explanatory: age, height, minutes played, and fouls committed. The remaining nine characteristics – pace, offensive efficiency, coaching and player continuity, athleticism, wingspan, experience prior to the NBA, total college players, and individual PER – are further defined below:

Once a team's total possessions have been calculated, pace is simply the average number of possessions a team uses during the course of a game. In the NBA, a typical team has a pace of around 100 possessions per game.

Defensive efficiency illustrates how many points a defense allows on a per-possession basis. Similarly, offensive efficiency illustrates how many points an offense scores on a per-possession basis. This study used points per 100 possessions for both defensive and offensive efficiency.

Coach continuity was defined as the number of years the head coach for each lineup was employed as the head coach of that organization. This variable assumes that a longer tenure allows a coach a better opportunity to implement an effective defensive system. While not all coaches have the same defensive acumen, Oliver's research shows that there is some correlation between poor defensive teams and recently hired head coaches.

Player continuity was defined as the number of years that player has been a member of the team. The overall player continuity of a lineup was calculated by finding the average player continuity of the five players in that specific lineup.

A lineup's athleticism was determined by their combined steal and block rate, which is simply the percentage of defensive possessions that end in a steal or block. While this metric has inherent correlations with the success of a defense, it has historically been used as one way to measure an individual's athleticism, especially in pre-draft analysis. This study also uses combined steal and block rates as a descriptor for a lineup's athleticism, albeit with a note of caution due to the nebulous nature of measuring athleticism.

Player height was taken from the in-shoe height listed at the Pre-Draft NBA Combine. If the Combine measurement was not available, the team-listed height was used. The wingspan length was also taken from their Combine measurements. If a player's wingspan was not available, their wingspan was represented by their height measurement.

Experience prior to the NBA measured the amount of experience gained through either college or professional leagues outside the NBA before entering the NBA. This combination accounts for all classifications of players regardless of whether they attended college, went straight to the NBA after high school, or played in non-NBA leagues such as those in Europe and China. However, in an attempt to gauge the effect that college basketball has on the NBA, each lineup's total number of college players was also calculated.

Player Efficiency Rating, commonly known as PER, is a statistic created in the 1990s by John Hollinger, Vice President of Basketball Operations with the Memphis Grizzlies. PER takes into account many different team and individual factors in an attempt to rate the performance of an individual player. In this study PER was used to estimate a player's overall basketball ability, despite the offensive leanings of the statistic. PER is based on a player's box score, which as previously explained contains more offensive metrics. As a result, offensive specialists typically have a higher PER than defensive specialists.

Significance

As an integral part of contemporary American culture, sports play a significant role in the lives of coaches and players as well as their families and fans. Today the field of advanced analytics and prediction has revolutionized the modern sport movement. Each time someone takes a critical look at a small piece of a game, another page is added to that game's playbook. As coaches, players, and fans study and apply the knowledge contained on each page, their knowledge and understanding of the overall game also increases. Coaches at all levels of any game – youth, college, or professional – can access and use the data and information gained from this study. In other words, by implementing this research and analysis of certain NBA team characteristics in terms of both roster construction and coaching, coaches will be able to better ascertain and estimate their team's defensive ability.

Several variables could help support several coaching strategies as to their defensive significance such as the team's pace, whether players should try to create turnovers, and how many minutes each player should be on the court. And while analytics may be an unusual way of looking at the defensive success of a team, it may prove significant as an added strategy for roster and lineup construction and a vehicle to rely on as sporting travelers explore the great unknown.

Findings

As in any journey in life, the effort and time spent on the trip is often just as rewarding as reaching the destination. I first collected and organized the data. I then overviewed the data, systematically reviewed the data, and finally analyzed the data. Along the way, a clearer picture and understanding of how identified variables affect team defense slowly emerged. The destination was now in view, that place where the relationship between roster construction, coaching strategy, and team defense are clearly apparent and understood.

As in any journey, there were stops along the way. The first stop was at a data overview where a sightseeing tour of the summary statistics for each dependent and independent variable investigated in this study was undertaken as shown in Table 1. There were 113 five-man lineups over the time span sample (2008-14) that, as a unit, played over 400 minutes in one season.

As seen in Table 1, the average lineup's Offensive Efficiency was over 5 points per 100 possessions higher than the average lineup's Defensive Efficiency. Thus, the average lineup sampled was better than the "theoretical average" lineup which would have equal Offensive and Defensive Efficiencies since offense and defense are two sides of the same coin. In other words, basketball is a zero-sum game where every point scored is a point allowed, just as every time the coin comes up heads, it does not come up tails. This finding makes sense logically, as only the best lineups would be allowed to play extended minutes over a season.

This idea that a subset of above-average players were sampled is further born out in the PER mean. The average PER for all players in the league is 15. However, the average of players sampled in this study is 16.63.

Another interesting observation is that the mean of coach continuity is greater than the mean of player continuity. The NBA is often described as a player's league where the players are the most important-decision makers in the league. It should then follow that players should have the greatest staying power with any given team. However, the lineups sampled show that coaches have greater staying power than players. This assumption does not take into account underlying factors relating to coaches such as a smaller population and fewer career limitations including injury, age, and finances that may apply to players.

Table 1

Summary Statistics of 14 Variables

n = 113	Mean	St. Deviation	Minimum	Maximum
Defensive Efficiency	101.953	4.709	89.1	112.7
Offensive Efficiency	107.415	4.738	95.0	117.1
Pace	94.420	3.512	86.78	102.85
Total Minutes Played	654.885	259.448	400.00	1468.00
Foul Rate	.184	.019	.139	.236
Age	26.678	2.325	21.2	32.4
PER	16.634	1.301	12.86	19.76
Average Height	79.070	1.400	75.565	82.400
Average Wingspan	81.479	1.915	76.74	85.55
Average Player Continuity	3.549	1.228	1.6	7.4
Coach Continuity	4.354	3.939	1	22.5
Steal and Block Rate	.135	.018	.089	.190
Average Prior to NBA Experience	2.660	.846	1.2	6
Total College Players	3.752	.987	1	5

On the second stop, a quick maintenance check was completed to examine the relationships between the independent variables. In order to determine if there were any variables that were collinear, that is if any two variables were highly correlated and could affect the results of a regression analysis, pair-wise correlations for each independent variable were calculated. As seen in Table 2, the highest correlation was .6856 which is below the generally accepted threshold for collinear variables of .9 as well as the highest correlation possible of 1.0. Therefore, the journey could continue and the analysis could proceed as originally planned.

A side-note on Table 2, the suspicions of PER having a bias towards offense appear to be well-founded as PER and Offensive Efficiency had the highest correlation among the variables tested with a correlation of .6856.

Table 2
Pair-Wise Correlations

	Off. Eff.	Pace	Total Minutes Played	Foul Rate	Age	PER	Avg. Height	Average Wingspan	Average Player Cont.	Coach Cont.	Steal and Block Rate	Average Prior to NBA Experience
Pace	.0788											
Total Minutes Played	.1219	-.0465										
Foul Rate	-.1036	.0467	-.0739									
Age	.3093	-.1801	-.0361	-.1467								
PER	.6856	.1399	.1995	-.1150	.3006							
Average Height	.0387	-.0123	.0211	-.1822	.0029	.0213						
Average Wingspan	-.2301	-.0796	-.0097	-.1740	-.3405	-.1449	.6223					
Average Player Continuity	.2929	-.1333	.1687	-.1002	.6545	.3844	-.0164	-.3101				
Coach Continuity	.1899	-.0728	.0260	.0345	.3120	.3194	.0333	-.1026	.5744			
Steal and Block Rate	-.0319	-.0623	.1207	.1157	.0522	.2340	-.1335	-.1030	.1320	.1150		
Average Prior to NBA Experience	-.0773	-.1000	-.0416	-.0779	-.1518	-.1738	.2244	.0831	.0057	.1189	-.2514	
Total College Players	-.2513	-.0647	-.0198	.0395	.1682	-.4123	.0784	.3926	-.2656	-.3000	-.0499	.0586

Once the dataset was collected, observed, and determined to have no collinearity issues, the first regression destination was reached. The first regression analysis performed was a full analysis with all 13 independent variables. A results chart from this analysis is included in Table 3.

Table 3
Full Model Results Chart

Results Chart: Y=Defensive Efficiency

Variable (Units)	Slope	p-value	Statistically Significant?	Interpretation
Offensive Efficiency (Points/100 Possessions)	.0116	.9217	No	
Pace (Possessions)	.2593	.0239	Strongly Significant	A team gives up .2593 points per 100 possessions for each possession used.
Total Minutes Played (Minutes)	-.0007	.6385	No	
Foul Rate (Fouls/Possession)	-18.5542	.3817	No	
Average Age (Years)	-.3226	.1853	No	
PER	-.1994	.6842	No	
Average Height (Inches)	-.2544	.5056	No	
Average Wingspan (Inches)	.3913	.2458	No	
Average Player Continuity (Years)	-.7655	.1285	Weakly Significant	A team gives up .7655 fewer points per 100 possessions for each year a five-man lineup stays with the same team.
Average Coach Continuity (Years)	.3109	.0136	Strongly Significant	A team gives up .3109 points per 100 possessions for each year a coach stays with the same team.
Steal and Block Rate (Steal or Block/Possession)	-108.8828	.0000	Very Strongly Sig.	A team gives up 1.089 fewer points per possession for each steal or block recorded in 100 possessions.
Average Prior to NBA Experience (Years)	-.8017	.1058	Weakly Significant	A team gives up .8017 fewer points per 100 possessions for each year a five-man lineup plays outside of the NBA.
Total College Players (Players)	-.1906	.6923	No	

Of the thirteen independent variables, only five were determined to be statistically significant. Thus, for the other eight variables, the initial hypothesis that there is no relationship between the variable and defensive efficiency is not rejected. That is, there was not enough evidence to say that there is a relationship between offensive efficiency and defensive efficiency, total minutes played and defensive efficiency, foul rate and defensive efficiency, age and defensive efficiency, PER and defensive efficiency, height and defensive efficiency, wingspan and defensive efficiency, or total college players and defensive efficiency. However, for the other five variables, the initial hypothesis was rejected and the variables of pace, player continuity, coach continuity, steal and block rate, and prior to NBA experience each has a relationship with defensive efficiency. This model has an r-squared value of .3183. Thus, these thirteen variables together explain about one-third of the variation of defensive efficiency.

In order to continue on the journey to reach the final destination, a backwards-elimination regression analysis was performed. A backwards elimination is a process where the most statistically insignificant variables are eliminated, one at a time, until only statistically important variables remain and each p-value is less than .15. A results chart for this backwards elimination, or final, regression model is provided in Table 4.

Table 4

Backwards Elimination Model Results Chart

Results Chart: Y=Defensive Efficiency

Variable (Units)	Slope	p-value	Statistically Significant?	Interpretation
Pace (Possessions)	.2207	.0381	Strongly Significant	A team gives up .2207 points per 100 possessions for each possession used.
Average Age (Years)	-.3841	.0749	Moderately Significant	A team gives up .3841 fewer points per 100 possessions for each year a five-man lineup ages.
Average Player Continuity (Years)	-.9137	.0515	Moderately Significant	A team gives up .9137 fewer points per 100 possessions for each year a five-man lineup stays with the same team.
Average Coach Continuity (Years)	.3181	.0136	Strongly Significant	A team gives up .3181 points per 100 possessions for each year a coach stays with the same team.
Steal and Block Rate (Steal or Block/Possession)	-115.3338	.0000	Very Strongly Sig.	A team gives up 1.153 fewer points per possession for each steal or block recorded in 100 possessions.
Average Prior to NBA Experience (Years)	-.7720	.0971	Moderately Significant	A team gives .7720 fewer points per 100 possessions for each year a five-man lineup plays outside of the NBA.

The final regression model revealed 6 statistically significant independent variables as opposed to just 5 in the full regression model. The additional variable was age which was determined to have a moderately significant p-value of .0749. This final model p-value for age is .1104 less than the full model p-value of .1853. Age was also determined to have a slope of -.3841. That is, a team gives up .3841 fewer points per 100 possessions for each year a five-man lineup ages.

However, the weakest statistically significant variable was prior to NBA experience with a moderately significant p-value of .0971. This final model p-value is .0087 less than the full model p-value. Prior to NBA experience was also determined to have a slope of -.7720. In other words, a team gives up .7720 fewer points per 100 possessions for each year a five-man lineup stays with the same team. This final model interpretation represents .0297 fewer points allowed than the full model.

The third weakest statistically significant variable was player continuity with a moderately significant p-value of .0515; this final model p-value is .077 less than the full model p-value. Player continuity was also determined to have a slope of -.9137. That is, a team gives up .9137 fewer points per 100 possessions for each year a five-man lineup stays with the same team. This final model interpretation represents .1472 more points allowed than the full model.

The third strongest statistically significant variable was pace with a strongly significant p-value of .0381; this final model p-value is .0142 more than the full model p-value. Pace was also determined to have a slope of .2207. That is, a team gives up .2207 points per 100 possessions for each possession used over the course of a game. This final model interpretation represents .0386 less points allowed than the full model.

The second strongest statistically significant variable was coach continuity with a strongly significant p-value of .0136. This final model p-value is .0071 less than the full model p-value. Coach continuity was also determined to have a slope of .3181. That is, a team gives up .3181 points per 100 possessions for each year a coach stays with the same team. This final model interpretation represents .0072 more points allowed than the full model.

The strongest statistically significant variable was steal and block rate with a very strongly significant p-value of less than .0001; this final model p-value is not significantly different than the full model p-value. Steal and block rate was also determined to have a slope of 115.33. That is, a team gives up 1.153 fewer points per possession for each steal or block recorded in 100 possessions. This final model interpretation represents .0645 more points per possession allowed than the full model.

The backwards elimination regression model had an r-squared value of .3423. In other words, these six variables together explain 34% of the variation found in defensive efficiency. Due to the presence of more variables as well as a higher r-squared value, the backwards elimination regression model was the main regression model used throughout the rest of the study.

Before continuing on to prediction, a more extensive maintenance check on the regression model's assumptions was conducted. First, a plot of the predicated values vs. the residuals was created (Appendix A). This plot showed a random pattern around the regression line. Thus, the variance of residuals was reasonably constant and that a linear model fit the dataset reasonably well. Next, a normal probability plot of the residuals was created (Appendix B). This plot was reasonably linear and, thus, the residuals were determined to be normal. The regression model also assumes that the residuals are independent, each data value is equally reliable, and the data values are measured without error. These final three assumptions can be made with a reliable sampling design, which was achieved in this study.

In order to determine which values were most damaging to the regression model, a Cook's Distance Plot (Appendix C) was also created and compared to the predicted values vs. residuals plot. The only data point found to have a large residual while also being highly influential was a 2012-13 Utah Jazz lineup. This lineup, consisting of Jamaal Tinsley, Randy Foye, Marvin Williams, Paul Millsap, and Al Jefferson, had a defensive efficiency of 108.6. Despite being in the worst 25% of defensive lineups sampled, this lineup was also one of the lowest 25% lineups in terms of pace and one the highest 25% lineups in terms of steal and block rate. Usually a lineup with these pace and steal and block rate numbers would result in a good defensive lineup.

The last hypothesis to be tested in this study was that the roster construction variables had a stronger relationship with defensive efficiency than the coaching strategy variables. In order to address this hypothesis, a regression analysis was conducted for each set of independent variables. A results chart for the roster construction variables is provided in Table 5, and a results chart for the coaching strategy variables is provided in Table 6.

Table 5
Roster Construction Results Chart

Results Chart: Y=Defensive Efficiency

Variable (Units)	Slope	p-value	Statistically Significant?	Interpretation
Average Age (Years)	-.3697	.1117	Weakly Significant	A team gives up .3697 fewer points per 100 possessions for each year a five-man lineup ages.
PER	.0133	.9698	No	
Average Height (Inches)	-.1551	.6791	No	
Average Wingspan (Inches)	.3004	.3346	No	
Average Player Continuity (Years)	-.8783	.0735	Moderately Significant	A team gives up .8783 fewer points per 100 possessions for each year a five-man lineup stays with the same team.
Average Coach Continuity (Years)	.3009	.0157	Strongly Significant	A team gives up .3009 points per 100 possessions for each year a coach stays with the same team.
Steal and Block Rate (Steal or Block/Possession)	-117.9719	.0000	Very Strongly Sig.	A team gives up 1.1791 fewer points per possession for each steal or block recorded in 100 possessions.
Average Prior to NBA Experience (Years)	-.8591	.0850	Moderately Significant	A team gives up .8591 fewer points per 100 possessions for each year a five-man lineup plays outside of the NBA.
Total College Players (Players)	-.1723	.7155	No	

Table 6
Coaching Strategy Results Chart

Results Chart: Y=Defensive Efficiency

Variable (Units)	Slope	p-value	Statistically Significant?	Interpretation
Offensive Efficiency (Points/100 Possessions)	-.0872	.3494	No	
Pace (Possessions)	.3477	.0059	Very Strongly Significant	A team gives up .3477 points per 100 possessions for each possession used.
Total Minutes Played (Minutes)	-.0019	.2582	No	
Foul Rate (Fouls/Possession)	-21.8977	.3495	No	

The roster construction regression model had a very strongly significant p-value of less than .0001 and an r-squared value of .3023. In other words, the variables determined to indicate roster construction explained about 30% of the variation found in defensive efficiency. The coaching strategy model had a strongly significant p-value of .03635 and an r-squared value of .05609. In other words, the variables determined to indicate coaching strategy explained only 5% of the variation found in defensive efficiency. Since the roster construction variables were more statistically significant and explained 25% more of the variation found in defensive efficiency, the initial hypothesis was confirmed.

Proposal/Recommendations

Before starting this journey, I identified a problem of the lack of knowledge regarding the importance of certain characteristics in the development of team defense in basketball. I then asked the question: How do certain team characteristics, in terms of both roster construction and coaching strategy, affect a team's defensive ability? Now, at the end of the journey, I propose that executives, coaches, and fans utilize the regression analyses as presented in this paper as a guide to better understand how given variables affect team defense. Simply put, I propose that these findings may help teams win basketball games.

I believe the results as reflected in the fitted regression model clearly show that these things do matter and should be taken into account. One useful time to take a closer look at the defensive ability of your team as well as opposing teams would be during the off season. Table 4, the Results Chart from the backwards elimination regression model, records the findings of six statistically significant variables. A better understanding of the pace, player age, player and coach continuity, as well as steal and block rate of your team will help executives make informed selections in player acquisition, will help coaches select the best style of play for various lineups, and will also help fans better understand what makes their team tick.

Another good time to take a look at these variables could be at mid-season. Knowing the defensive ability of potential trade targets as well as tradeable team members would help executives make effective trade decisions. Likewise, coaches could use the stats from the first half of the season to evaluate the team's strengths and weaknesses and take appropriate steps to continue its winning ways or help turn the team around. Fans, especially, could benefit from a realistic look at their team at mid-season as they look ahead to either post-season play or next year.

Perhaps the first question that most basketball enthusiasts would ask would be, "But what about during the game itself?" While there are meaningful coaching strategies that can be implemented during a game that matter, this study did not examine variables pertaining to the product on the court, except when no other viable option was available as in the case of determining athleticism. A decision was specifically made to look at team-makeup factors that could be measured prior to game action. As a result, this model confirmed the hypothesis that roster construction is more important than the coaching strategy, at least as defined in this paper.

Expanded Discussion of Significance

As an integral part of contemporary American culture, many people play, support, or follow sports. Today the field of advanced analytics and prediction has become as important and crowded as the playing field. In addition to executives, coaches, and fans of the NBA, people involved at all levels of the game of basketball can benefit from the general concepts and findings of this study. While analytics may be an unusual way of looking at the defensive success of a team, I believe it is a relevant and reliable strategy that adds a new page in the playbook of the game of basketball.

Following are just a few of the many examples wherein the fitted regression model might help provide an answer. When looking to fill an NBA roster, one variable that should be taken into account is the player's steal and block rate which is a very strongly significant factor in a team's overall defensive ability. The model showed a team gives up 1.153 fewer points per possession for each steal or block recorded in 100 possessions. That's enough to win the game.

Most basketball fans enjoy watching a face-paced, up-tempo game. In other words, fans want to be entertained. However, this study concluded that if a team plays at a higher pace, they will give up more points because their rate of points allowed also goes up. This finding supports the research of Dean Oliver (2004) who found that the best defensive teams averaged one-half fewer possessions per game than the average team. In other words, slower is better as far as defensive ability and winning games through good defense.

However, there was one finding that ran counter-intuitive to Dean Oliver's research. This study concluded that the longer a coach stays with the same team, the worse the team defense actually becomes. Dean Oliver showed that the worst defenses throughout history have had new coaches. Obviously, there is some nuance required in comparing and contrasting these results. Perhaps there is an unseen tipping point over time where coaches lose sway over their players and the defense depreciates, or perhaps defensive coaches do not play an entertaining enough style to be rewarded with long tenures.

Another significant factor is average player age. Many fans want to see the rookies and youngsters get more playing time. But as this study reflects, older veteran players make better defensive teams. In the same vein, players who have played on the same team for a number of years play better defense than newly-assembled lineups. So if you want to win games with defense, go with the veterans who have played together.

One factor with significance in today's NBA landscape is prior basketball experience outside the NBA through other professional leagues or college. Currently you must be at least 19 years old to play in the NBA. But while many superstars entered the NBA straight from high school or after only one year of college, this study showed that experience outside of the NBA has a positive relationship to defensive ability.

But perhaps the most significant application of this study's findings can be found in the realm of prediction. After all, one reason we study the past is to better prepare for the future. In this study, several archetypical predictions were derived from the backwards elimination regression model and the available dataset.

First, a “Worst Team” was created. This team would consist of the observation that most negatively affects team defense for each of the six characteristics as described in the final model, Table 4. In each of the following cases, defensive efficiency is presented in terms of points allowed per 100 possessions. This hypothetical worst team had a predicted defensive efficiency of 119.75 ±6. In other words, we are 95% confident that this team’s defensive efficiency would fall between 113.74 and 125.76 with a best guess of 119.75. We then created a “Best Team” consisting of the observation that most positively affects team defense for each of the six characteristics in Table 4. This hypothetical best team had a predicted defensive efficiency of 84.52 ±6. In other words, we are 95% confident that this team’s defensive efficiency would fall between 78.99 and 90.05 with a best guess of 84.52. Lastly, we created an “Average Team” consisting of the average observations in Table 4. This hypothetical average team had a predicted defensive efficiency rating of 101.67 ±1. In other words, we are 95% confident that the average team’s defensive efficiency would fall between 100.91 and 102.42 with a best guess of 101.67. These hypothetical lineup predictions are shown in Table 7.

Table 7
Hypothetical Lineup Predictions

Lineup Type	Prediction	95% Confidence Interval
Worst	119.75	113.74 to 125.76
Best	84.52	78.99 to 90.05
Average	101.67	100.91 to 102.42

We then chose seven lineups from the 2013-14 dataset whose players were on the same team at the start of the current season. This selection allows us to assume that each lineup would maintain its pace as well as block and steal rate into the upcoming season. We then adjusted the player age, player continuity, and coach continuity to reflect the new season.

The basketball experience prior to NBA data remained the same. Of course, team lineups are constantly subject to change due to uncontrollable factors such as injuries, coaching changes, and player transactions that affect the makeup of the team's lineup and who plays on any given night. These results are shown in Table 8.

In order to make meaningful comparisons, we then specifically looked at two of these teams, the Portland Trail Blazers and Toronto Raptors, whose lineups have remained relatively stable. Looking at Portland, we predict a defensive efficiency of 105.16 and are 95% confident that this lineup's defensive efficiency would fall between 102.39 and 107.92. As of December 7, 2014, this particular Portland lineup has a defensive efficiency of 98.2, which reflects a better defense than our prediction. However, this lineup has also played at a much slower pace than assumed which would account for some of the discrepancy. In 2013-14 Portland had one of the fastest paces in the dataset while currently they are playing closer to the dataset's average. In Toronto's case, we predict a defensive efficiency of 104.93 and are 95% confident that this lineup's defensive efficiency would fall between 103.11 and 106.75. As of December 7, 2014, this particular Toronto lineup had produced a defensive efficiency of 102.0, which again represents better defensive efficiency than predicted. In Toronto's case, however, there is not a readily apparent reason why they are playing better defense than predicted.

Table 8
Specific Team Predictions

Team	Prediction	95% Confidence Interval	Lineup Production	Team Production
Portland	105.16	102.39 to 107.92	98.2 – 319 min	99.5
Golden State	100.75	99.15 to 102.35	0 min	95.1
Toronto	104.93	103.11 to 106.75	102.0 – 213 min	104.0
Memphis	98.47	96.42 to 100.52	107.6 – 44 min	99.4
Denver	104.00	102.20 to 105.79	0 min	105.7
Detroit 1	103.94	102.24 to 105.63	104.5 – 111 min	104.5
Detroit 2	104.66	103.28 to 106.04	50.0 – 2 min	

If one is willing to make a few assumptions while leaping from a micro to a macro view, these specific lineup predictions may also be used to predict a team's overall defensive production. As of December 7, 2014, four of the six teams represented in the predictions have a team defensive efficiency within the bounds of what we would expect. The other two teams are outperforming expectations with lower defensive efficiencies than predicted.

Summary/Closure

Looking back over the journey of this research project, I am reminded of the problem – a lack of knowledge and research on the defensive side of basketball – as well as the purpose of the trip – to provide executives, coaches, and fans with a clearer picture and understanding of how given identified variables affect team defense.

Throughout the process, I have enjoyed the trip. On the first leg I learned about the process of collecting and organizing large amounts of data. One of the challenges I encountered was the imperfect nature of raw data. In identifying the variables, athleticism was an especially difficult concept to define and quantify. My first idea was to use individual quickness drill data from the NBA Draft Combine. However, very few of the players in this study participated in the quickness drills. My second idea was to use the individual steal and block rates from each player's collegiate career. Again, steal and block rate stats were either not available or reliable for many of the players. I ended up creating a team steal and block rate from the given data on the NBA Stats website.

On the second leg of the trip, I learned the finer points of regression analysis. And while the final results are useful, they could have been better. After all, the answer we get depends on the question we ask, and there just is not a good answer for some questions. Sometimes we can only find what the data tells us, and we do not always get the results we're hoping for.

On the final leg I learned how to better present the statistical findings and their applications. In-depth analysis of a subject requires living and breathing that project for an extended period of time. While not an easily acquired skill, I am steadily learning how to step back from the minutiae to present overarching themes and findings.

The journey has ended, the destination reached, and the vacation scrapbook completed. Hopefully the findings presented in this paper will give people a better appreciation and understanding of various defensive variables as found in the NBA. Perhaps the process of identifying characteristics and examining their relationship to the final outcome can be applied in other arenas of life as well. I look forward to the next trip into the unknown.

References

Ayer, R. (2012). Big 2's and Big 3's: Analyzing how a team's best players complement each other. *MIT Sloan Sports Analytics Conference.*

Cervone, D., D'Amour, A., Bornn, L., & Goldsberry, K. (2014). Pointwise: Predicting points and valuing decisions in real time with NBA optical tracking data. *MIT Sloan Sports Analytics Conference.*

Ergül, B., Yavuz, A. A., & Yavuz, H. S. (2014). Classification of NBA league teams using discriminant and logistic regression analyses. Pamukkale Journal of Sport Sciences, 5(1), 48-60.

Franks, A., Miller, A., Bornn, L., & Goldsberry, K. (2014). Characterizing the spatial structure of defensive skill in professional basketball. Manuscript submitted for publication.

Gems, G. R., Borish L.J., & Pfister, G. (2008). *Sports in American History: From colonization to globalization.* Champaign, IL: Human Kinetics

Goldsberry, K. (2012). CourtVision: New visual and spatial analytics for the NBA. *MIT Sloan Sports Analytics Conference.*

Goldsberry, K., & Weiss, E. (2013). The Dwight effect. *MIT Sloan Sports Analytics Conference.*

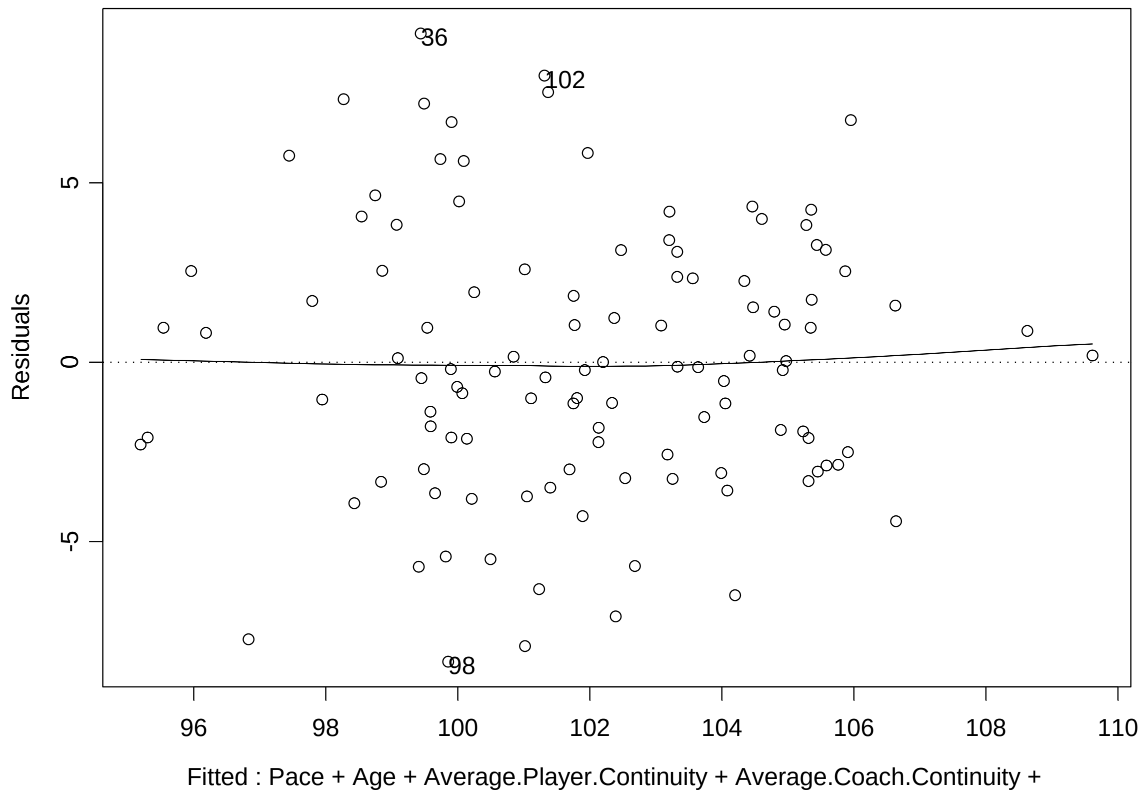
Maheswaran, R., Chang, Y., Su, J., Kwok, S., Levy, T., Wexler, A., & Hollingsworth, N. (2014). The three dimensions of rebounding. *MIT Sloan Sports Analytics Conference.*

Oliver, D. (2004). *Basketball on Paper: Rules and tools for performance analysis.* Washington, DC: Potomac Books

Sampaio, J., Ibáñez, S., Lorenzo, A., & Gomez, M. (2006). Discriminative game-related statistics between basketball starters and nonstarters when related to team quality and game outcome. Perceptual & Motor Skills, 103(2), 486-494.

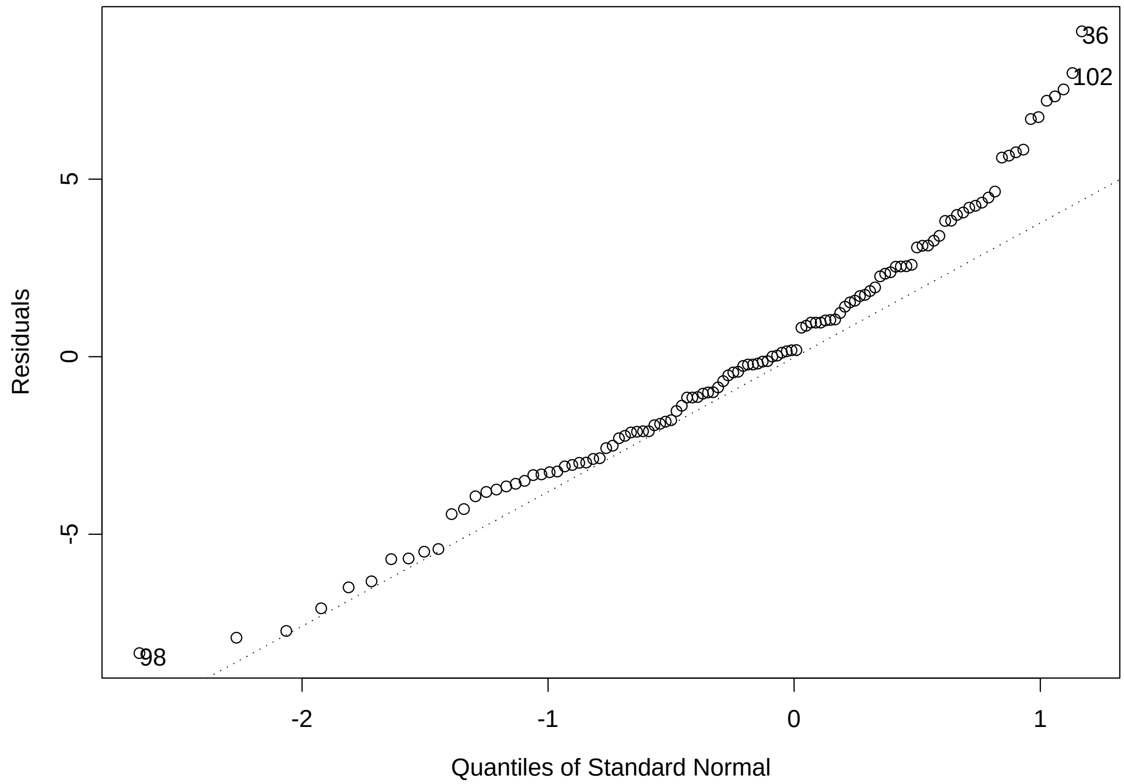
Appendix A

Residuals vs. Predicated Values Plot



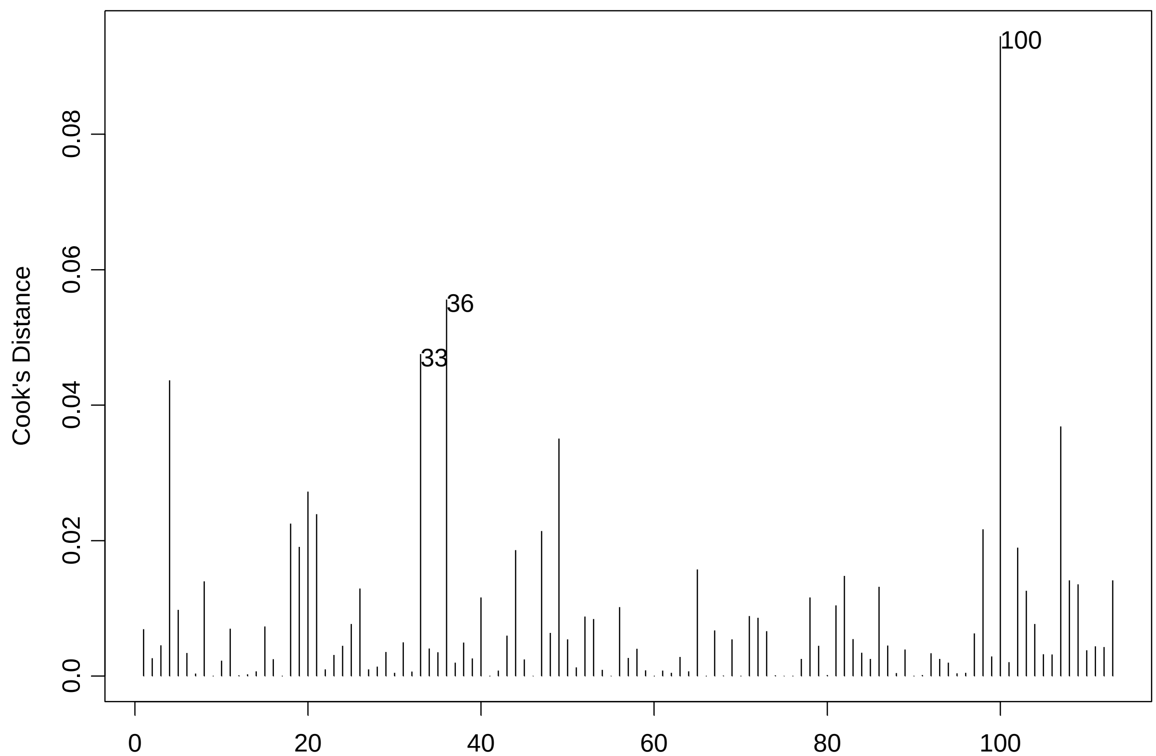
Appendix B

Normal Probability Plot of Residuals



Appendix C

Cook's Distance Plot



This Study by: Luke Peterson

Entitled: In Defense of Defense: A Statistical Look at Roster Construction, Coaching Strategy, and Team Defense in the National Basketball Association

has been approved as meeting the thesis or project requirement for the Designation

University Honors with Distinction

12/16/14
Date

Dr. Mark Ecker, Honors Thesis Advisor

12/19/14
Date

Dr. Jessica Moon, Director, University Honors Program