Spring 2018

Win Shares & Rookie Contracts in the NBA

Lucas R. Kobat
University of South Dakota

WIN SHARES & ROOKIE CONTRACTS IN THE NBA

by

Lucas R. Kobat

A Thesis Submitted in Partial Fulfillment
Of the Requirements for the
University Honors Program

Department of Economics
The University of South Dakota
May 2018

The members of the Honors Thesis Committee appointed

to examine the thesis of Lucas R. Kobat

find it satisfactory and recommend that it be accepted.

Dr. Mike Allgrunn
Professor of Economics
Director of the Committee

Ms. Mandie Weinandt
Instructor of Economics

Mr. Tyler Custis, J.D. & MBA
Lecturer of Business Law

ABSTRACT

Win Shares & Rookie Contracts in the NBA

Lucas R. Kobat

Director: Mike Allgrunn, Ph.D.

This paper examines the relationship between the win share statistic and compensation in the National Basketball Association (NBA) by using data from eighteen draft classes from 1989 to 2006. The research shows that players are generally compensated in accordance with their production, unless bound by a rookie contract. Historically, players under a rookie contract have win share production that exceeds their compensation level. Therefore, in-game statistics are examined, using both collegiate and NBA data to determine whether win share production can be predicted before a player enters the NBA. Collegiate data does not prove to be a sound indicator of professional level win-shares, but in-game statistics do seem to be a good predictor of win-shares when NBA data is used. Ultimately, win share regression is beneficial for NBA organizations making rookie contract decisions (i.e. team options) for players that have been drafted, but further research would be needed to determine which players to draft.

KEYWORDS: Win Shares, Rookie Contracts, NBA, Advanced Metrics, Sports

TABLE OF CONTENTS

Table of Figures

Preface

This paper seeks to use regression analysis and other statistical methods to understand the relationship between the individual contribution of players in the National Basketball Association (NBA) and their compensation. Ultimately, the research aims to develop a methodology for rookie contract decisions. Organizations take risk when they draft players, and risk/reward analysis must be completed to assess whether or not it is beneficial for organizations to exercise team options and continue to compensate prior draft selections.

NBA organizations, at the end of the day, are firms that care about profit margins. While much of the literature attempts to capture marginal revenue per player through various methods, this research discusses individual player contribution as it relates to wins, disregarding intangible factors that may indirectly impact team revenues, such as personal brand. Forecasting brand affinity for a particular player is less reliable due to myriad variables that lie outside the realm of basketball, whereas win contribution is a direct impact to the success or failure of a player's respective team.

CHAPTER ONE

Introduction

A Brief Argument For the Importance of Wins

The research in this paper hinges on the assumption that additional wins are the most economically beneficial variable to organizations in the NBA. Therefore, additional wins will be viewed as the root cause of success for NBA teams. Essentially, more regular season wins result in an increased chance of winning an NBA championship; increasing team monetary value.

Over the time period of 1989 to 2006, teams that either had the most wins in the Eastern Conference or Western Conference won the NBA championship 56% of the time. For championship winners, the average number of wins over this time period was sixty. Sixty wins equates to roughly 73% of games played in a regular season, as every team plays eighty-two games each year. Since teams need to win 73% of their regular season games to win an NBA championship, wins seem to be highly important to team success. Why do teams care about winning championships? The impact of championships on both short-term and long-term growth of team monetary value is substantial and shows why teams should care about winning the NBA championship. For example, the winner of the NBA championship in 2017 was the Golden State Warriors. Their current value is $2.6 billion, but the important statistic to note is the 37% one-year growth they experienced after winning the championship. In fact, this growth rate was 1.31 times more than the second highest growth rate at 16%. This short-term growth is not atypical of NBA champions. The Cleveland Cavaliers, after winning their first championship in franchise history, grew their team monetary value by 78% from $515 million in 2014 to $915 million in 2015.

Long-term growth also accompanies teams that are perennially successful in the post-season. As of 2017, four of the top five most valuable teams were also in the group of five teams with the most championships wins. In fact, the New York Knicks were the only team to be in the top five most valuable, but not the top five most championship wins. However, if NBA championship appearances were used in lieu of championship wins, all top five most valuable teams would also be in the group of five teams with most NBA finals appearances.

As has been stated, regular season wins are a strong determinant of both NBA championship winners and teams that make the NBA finals. Additionally, teams that win the NBA championship experience a substantial short-term growth in monetary value, and teams that have historically been successful in the post-season experience long-term growth in monetary value. Therefore, NBA organizations should be interested in finding players that positively impact their yearly win total. To maximize their probability of growing monetary value, NBA teams should not only locate players with positive win contributions, but also players that do not carry a heavy payroll expense.

This paper sought to find players with positive win contributions and minimal payroll burdens. Finally, the research aimed to create a comprehensive methodology for predicting win contribution levels both pre-draft and post-draft, which can aide NBA organizations in rookie contract decisions.

CHAPTER TWO

Literature Review

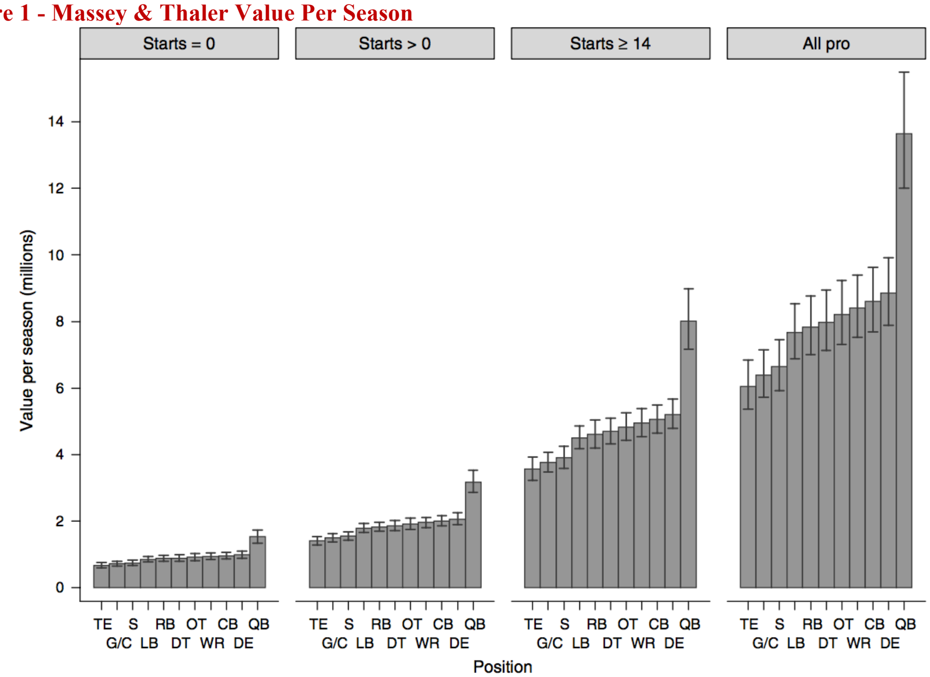
Dr. Thaler & Dr. Massey – The NFL Draft

The inspiration for this research was largely based on the work of Dr. Richard Thaler of the University of Chicago and Dr. Cade Massey of the University of Pennsylvania. The work of Dr. Thaler and Dr. Massey involved studying the efficient market hypothesis in the National Football League (NFL).

The duo used their data set to examine value of NFL players over their career. They used certain variables to capture "value" including the amount of starts, Pro Bowl selections, and yearly compensation. Player positions were also used in the performance evaluation (Massey & Thaler, 2013).

Figure 1 illustrates value per season for all player positions measured in millions of US dollars:

Figure 1 - Massey & Thaler Value Per Season



Notes. The labor market value of a player's previous five years. These are the predicted values from model (2) in Table 2, estimated from compensation in player years 6–8 for draft classes 1991–2001. Note that these are values for player performance that falls into a category 100% of the five-year history. Reported in 2008 dollars.

Clearly, quarterbacks are more valuable in the NFL marketplace than other positions, and increasingly more so as they become more successful. For example, a quarterback who starts fourteen or more games in one season is worth roughly $8 million dollars. However, if that quarterback is also "All pro", their value jumps to just shy of $14 million dollars, which is roughly a 75% increase.

This examination of value is precisely the research this paper seeks to conduct, but through win contribution levels, taking the above method a step further by examining individual performance and the correlation to team wins generated. This expansion of the work done by Dr. Massey and Dr. Thaler allows for varying monetary values to be assigned to players according to their predicted win contribution.

An Explanation of Win Shares

Basketball is a fluid, fast-paced game with five individuals executing a variety of tasks simultaneously to produce the end product of points. This paper seeks to assign win-contribution levels to individual players. The concept of individual win-contribution is not completely new, and several sports sabermetricians have developed their own models.

The first to ever develop a win-contribution, or win share, model was Bill James. James, who currently works for the Boston Red Sox as Senior Baseball Operations Advisor, began his work on advanced baseball statistics in the 1970's. Known as the pioneer of the sabermetric field, James became better known for *The Bill James Baseball Abstract,* in which he discussed the concept of win shares, along with many other ways of modeling Major League Baseball performance. His model assigns three win shares to every team per one win they produce, most likely to make the results easier to comprehend as most players (at least in the NBA) have sub-one win shares per season. The model works on an individual level as well as a team level. This paper is interested in win shares produced at the individual level, but for example purposes the model will be examined at the team level.

The Chicago Cubs in 2004 won 89 games. For Bill James' model to be accurate, the output should roughly equate to 89 games. The underlying mathematics of the model deal with marginal runs scored and marginal runs saved. The margin is defined as half of the league average runs scored and half the league runs allowed by batters and pitchers respectively. To calculate marginal runs scored for the Cubs, one must take the league margin in 2004 of 376 runs scored and subtract it from the runs scored by Chicago (789).

Then take the difference of the league margin in 2004 for runs allowed (1127) and runs allowed by Chicago (665). Both of these calculations work out to 413 hitting runs and 462 pitching runs. The combination of these figures equals 875 marginal runs for the Chicago Cubs in 2004. The 875 marginal runs are divided by twice the amount of league average runs in 2004 (1502), which produces an expected win-percentage of 58.2% or 94 wins. Since the 2004 Chicago Cubs won 89 games, the error is 1.05%.

The Sports-Reference Model

Workers at Sports-Reference.com, who focus on compiling data for professional and collegiate sports leagues, developed the win-share model used for this research. The model is specific to the NBA and deviates from the model developed by Bill James in two ways. First, it sets one win share equal to one team win, whereas Bill James' model set one win equal to three win shares. Second, the Sports-Reference model allows for negative win shares, which was avoided in Bill James' model with the lowest possible win share figure being zero. These two key distinctions improve the win share model by allowing direct comparisons of win contribution levels of individuals on a given roster and the team's win total. It also more realistically captures the notion that a player's performance could indeed hurt team performance to the extent that his win-contribution inhibits the win shares generated by his teammates.

Other NBA Prediction Models

This paper recognizes the magnitude of importance for the topic it examines, as the NBA is comprised of billion dollar organizations that have million dollar payrolls. Therefore, this research is not a final solution to predicting NBA player performance.

However, it does differentiate itself by taking a more focused approach in addressing the team option dilemma found within rookie contracts.

FiveThirtyEight's CARMELO

Most recently, writers at statistics publication *FiveThirtyEight* attempted to create a career projection tool named CARMELO. This model took player ratings like true shooting percentage and plus/minus scores, along with WAR (wins above replacement) to find similar historical players to the player being analyzed. Then, using the historical data of players most similar to the player in question, a ranged projected performance was calculated for statistics of interest like minutes played and WAR.

While the depth of the study is interesting, the breadth is limited with CARMELO ratings available solely for players who played at least 100 minutes in the 2014-2015 season or at least 250 minutes in the 2013-2014 season. In contrast, the research for this paper culminated statistics spanning eighteen draft classes from 1989 to 2006.

In terms of predicting win production, the CARMELO model only projects a ranged WAR based on similar players. First, WAR is valuable as a comparison tool, but not as an absolute figure. The statistic cannot cleanly be converted into actual wins; therefore it is not helpful in discerning individual win contribution. While WAR may be helpful when making certain roster decisions, win shares are most useful when attempting to predict win production levels because the statistic is tied to actual win results and not an arbitrary baseline.

Finally, as stated by the developers of the model, there is bias in the simulation as CARMELO favors certain players and does not have adequate data to compare others, such as Golden State Warriors standout Steph Curry. Curry's style of play is so unique that historical comparisons are merely a poor man's version of his performance, thus underweighting career projections (Silver & Paine, 2015).

Coates & Oguntimein Model

Coates & Oguntimein examined the effects of college statistics on NBA draft position. Their paper, published in the *North American Association of Sports Economists,* specifically looked at in-game collegiate statistics and whether the player was in a big or small conference during college. The duo then looked at the correlation between college performance and NBA performance at the career level and bifurcated the data by big and small conference players.

Their paper found that college statistics can be an indicator of NBA career success and that NBA organizations tend to commit more to early picks than late picks, which is possibly due to "escalation of commitment" bias. This discovery is important insight into team decisions regarding rookie contracts because it shows that players drafted earlier may receive undeserved contract extensions. For example, Kwame Brown (considered one of the major busts in draft history) kept his spot on the Washington Wizards roster for the first four years of his career. However, Coates & Oguntimein do not evaluate production at the individual level, nor do they examine whether NBA teams could solve the escalation of commitment issue by attempting to predict win share levels for team option years of the rookie contract (Coates & Oguntimein, 2008).

Greene's PER & Win Share Analysis

Alexander Greene, while conducting graduate research at St. Cloud State, published a paper quite similar to the research found in this thesis. Greene examined the effect of college statistics on draft position and rookie statistics on career performance.

The research focused on PER (Player Efficiency Rating) and win shares as end points of the study. As noted in Greene's paper, PER is not the most well rounded statistic, as it does not concretely attribute wins to individual players. Greene concluded that college statistics were good indicators of draft position and that rookie PER and rookie win shares were sound indicators of career performance.

Greene also used a similar data set, incorporating draft classes from 1985 to 2005 so that players drafted in the latter portion of the data set would have full career observations. Whether or not drafts pre-1989 should be intermingled with drafts post-1989 is a possible discussion point as 1989 was the implementation year for the modern draft lottery system. Additionally, Greene does not examine the nuance of the rookie contract and the team options that occur in years three and four. While taking similar initial steps, Greene's *"The Success of NBA Draft Picks: Can College Careers Predict NBA Winners?"* does not extend to the decision point faced by NBA organizations, which is simply whether or not a player should be kept after the second year of his rookie contract (Greene, 2015).

The NBA Draft

NBA Draft

The NBA draft dates back to 1947. The event encompasses the selection of college and foreign prospects by NBA franchises. Considerable amounts of resources are allocated to scouting incoming talent during their college or foreign careers. The modern NBA draft began in 1989 and reformed the process to include two rounds with a fixed draft order. NBA teams have the opportunity to add new talent to their respective rosters through the draft, which is one of only three opportunities to accomplish such a feat. The other two opportunities are through trades and free agency, both of which are more capital intensive than the draft (NBA Draft Rules, 2017).

NBA Draft Lottery

The first fourteen picks of the NBA draft are referred to as the lottery. Rights to draft at these positions are determined by a weighted Ping-Pong ball selection. Methodology relating to the amount of the weights and selection process has varied over time, but the framework has been consistent. Whichever team recorded the worst record the previous season is given the greatest weight in the lottery (currently measured by the count of Ping-Pong balls in the selection pool). The earliest versions of the draft lottery originate in 1985 with the modern system being implemented for the 1989-1990 season. The importance of the draft lottery is seen in trade behavior by NBA front offices that regularly attempt to "move up" into the lottery section of the draft in order to acquire better talent compared to the talent available in the remaining pool of players (Dengate, 2005).

NBA Rookie Contracts

Rookie contracts are two years in length, with a team-based option for a third year that may be exercised at the end of a player's rookie season until the following October 31st. If the team exercises their option for a third year, they are also entitled to an option for a fourth year. This option may be exercised from the completion of a player's second season to the following October 31st. Following the potential four years of a rookie contract, teams may extend qualifying offers to rookies, who at this point are considered restricted free agents (Jessop, 2012). Essentially, this restriction means that a player's current team has first mover advantage in offering a long-term contract.

Agents have little room to negotiate rookie contracts as they adhere to a strict salary scale. Compensation is dictated by the position a player is selected in the draft. However, under the current collective bargaining agreement (CBA) rookie salaries can fall within 80-120% of what the scale dictates. Typically, players selected in the first round (certainly within the lottery) command the 20% premium (Jessop, 2012).

Joshi Analysis

Nikhil Joshi, during his time at Stanford University, conducted research to determine whether top draft picks were overpaid compared to players drafted later. He concluded that rookies are roughly paid one quarter of what they would be worth in an open free-agent market. While not beneficial to the player, it does open the opportunity for NBA organizations to maximize production per payroll costs by performing well in the draft. Joshi attempts to capture marginal revenue per player to determine value, whereas this paper uses win shares that are dictated by actual in-game performance. Marginal revenue per player is a method that is widely used in the literature for comparing compensation, but it does not do a good job of determining when players are undervalued, which is why this paper chose to use win shares as its method of determining player value (Joshi, 2011).

Scouting Dilemmas

As has been discussed, the NBA draft is an opportunity for teams to add talent to their rosters without spending large amounts of capital. Due to the importance of this opportunity, NBA organizations place great emphasis on gauging the talent of incoming prospects through scouting. Yet, this method is not a perfect science.

One way NBA teams learn about incoming talent is through an organized event called the NBA Combine. This combine sheds light on a prospect's ability by measuring certain factors that effect performance such as bench press, vertical jump, and three quarter court sprint time. Additionally, physical characteristics like height, weight, and wingspan are recorded during the combine (Wasserman, 2017).

Teams also hire scouts to watch live and recorded performances of college and foreign prospects. Scouts usually come to a general consensus on players that will have the greatest impact in the NBA. However, this process has fallen victim to blatant misses over the years. Michael Jordan, considered to be one of the best players of all time, was drafted third overall in 1984 behind Hakeem Olajuwon and Sam Bowie. Olajuwon, although a two-time NBA champion, only found post-season success during the timespan that Jordan was away from basketball, pursuing baseball ambitions (*Biography.com*, 2016). Bowie, considered one of the greatest draft busts in NBA history, never won an NBA championship (Schoenfield, 1996).

A more recent and glaring example would be Draymond Green. Green is considered to be the prototypical "big-man" in the current small-ball era of the NBA. However, in 2012 Green fell to the 35th pick of the second round. There is no evidence to suggest that Draymond's game transcended significantly from the end of his college career to 2015, when Green was a main contributor to the Golden State Warrior's championship run. In hindsight, experts agree that the sole reason Green was not drafted higher was because scouts could not decide what position he was best suited for in the NBA (Titus, 2017). This scouting bias is the exact error that this research aims to help alleviate through analysis of advanced metrics.

Rating Agencies

Currently, rating agencies (as exist for high school prospects) are non-existent for college basketball players. Firms like 247, Scout and Rivals all have proprietary algorithms that attempt to measure high school prospect ability. For the NBA, media outlets like ESPN have analysts and senior writers that attempt to predict where a player will be drafted, but detailed analysis on a player's ability is relatively non-existent (Nusser, 2013). This lack of information detracts from a perfect market existing for the NBA draft and is possibly a cause for the errors mentioned in the previous section. Interestingly, 247 even have a "Top247" for college football players but do not have similar analysis for basketball. This gap in the rating agency market could possibly be an opportunity for one of the top firms, or a new firm who addresses this unmet need.

CHAPTER THREE

Methodology

Research Overview

The genesis of this research was the "Preliminary Draft Analysis" where the relationship between win-shares and draft position was examined. The "Preliminary Draft Analysis" included a data set of fourteen draft classes and was compiled using *Basketball-Reference.com*. The data set was expanded to include eighteen draft classes spanning 1989 to 2006 for the "Career Analysis". This analysis took career data for multiple variables including compensation and ran regressions against win shares in the attempt to find key drivers of individual win contribution. From this expanded data set a random sample of 35 players was taken to form a panel set. This “Panel Analysis” allowed for the examination of different variables’ effects on win shares over the course of a player’s career. From this analysis it became evident that the draft was of upmost importance, so all observations at the first, fifteenth, thirtieth, forty-fifth and sixtieth draft positions were taken from the master data set to examine the relationship between win share production over time at varying draft spots.

Equations

The Win Share Model

Mentioned previously in the “Sports-Reference Model” section, win share methodology was used for this research, in lieu of the Bill James model, due to the additional complexity that allows for negative wins and encompasses more statistics that are relevant to the game of basketball (NBA Win Shares).

To cover the Sports-Reference model in more detail, one must examine the two formulas that feed the final result: Offensive Win Shares (OWS) and Defensive Win Shares (DWS). OWS are calculated using Dean Oliver’s formulas for points produced and offensive possessions. All of the following formulas can be found in the Sports-Reference Glossary (*Basketball-Reference.com*). The step-by-step method is outlined below:

1) Calculate “Points Produced” for each player using Dean Oliver’s formula:

   a) $PointsProduced =$

   $$(PartialFG + PartialA\square T + FTMade) \times \left(1 - \frac{TeamOffRB}{TeamScoringPoss}\right) \times$$

   $$TeamOffRBWeight \times TeamPlay\%) + PartialOffRB$$

2) Calculate “Offensive Possessions” for each player using Dean Oliver’s formula:

a) $OffPoss = ScoringPossessions + FG\times Poss + FT\times Poss + Turnovers$

3) Calculate “Marginal Offense” for each player:

a) $MarginalOff =$

$(PointsProduced) - 0.92\times(LeaguePointsperPossession)\times(OffPoss)$

4) Calculate “Marginal Points per Win” for each player.

a) $MarginalPointsperWin = 0.32\times(LeagePointsperGame)\times(\frac{TeamPace}{LeaguePace})$

i) Where $TeamPace = 48\times(\frac{TeamPoss+OppPoss}{2\times(\frac{TeamMinPlayed}{5})})$

5) Calculate “OWS” for each player:

a) $OffWS = MarginalOff/MarginalPointsperWin$

Next, DWS must be calculated and added, as the sum of OWS and DWS results in the win share total for each player. This method is based partially on Dean Oliver’s “Defensive Rating” statistic, which is an estimate of a player’s points allowed per 100 defensive possessions. The step-by-step method is outlined below:

1) Calculate the “Defensive Rating” for each player using Dean Oliver’s formula:

a) $DefRtg = TeamDefRtg + 0.2\times(100\times DefPointsperScoringPoss\times(1 - Stop\%) - TeamDefRtg)$

2) Calculate “Marginal Defense” for each player:

a) $MarginalDef =$

$\left(\frac{PlayerMinPlayed}{TeamMinPlayed}\right)\times TeamDefPoss\times(1.08\times LeaguePointsperPoss) - (\frac{DefRtg}{100})$

3) Calculate “Marginal Points per Win” for each player:

a) $MarginalPointsperWin = 0.32\times LeaguePointsperGame\times(\frac{TeamPace}{LeaguePace})$

4) Calculate "DWS" for each player:

a) $DefWS = MarginalDef / MarginalPointsperWin$

The sum of OWS and DWS can then be calculated to arrive at a win share total for the respective player in question for any particular season.

WS/Season per Million Dollars

Wins-per-Million Dollars is an essential formula to this paper because it allows one to analyze the relationship between individual player contribution and compensation. As was stated in the "Basis of Research", one goal of this paper is to identify players with high win-contribution levels with minimal payroll burdens. To compute, the win shares produced for a specific time period is divided by the compensation accrued over the same time period, and then multiplied by one million. This equation is unique to this paper and is an extension of the formula list provided above.

$$\frac{WS}{Season} per\$1M = WS/Season \times \$1M \tag{1}$$

CHAPTER FOUR

Career Analysis

Description of Purpose

The overarching goal of this paper was to create a risk/reward methodology that aided NBA franchises in making rookie contract decisions. The first step in achieving the stated goal was to identify the types of players who historically have been high win share contributors and determine whether there are shared characteristics between them. The career analysis set out to determine key drivers of win shares by examining multiple variables' effect on player win-contribution, using career data. By expanding the variables from solely focusing on draft position, this step expanded the scope of the research and broadened its implications.

Data Summary

The data set for players drafted in the NBA between 1989 and 2006 originally consisted of 1,022 individuals. However, for the purposes of this research, the data set was parsed to 674 individuals who had at least played 82 games. This was done because when examining win shares over a specific period of time, such as wins per season or wins per 48 minutes, players who have only played a limited amount of games can skew the statistic. Using this method, only players who had played an equivalent of one season were included in the analysis. Of the 674 players, there were 120 point guards, 131 shooting guards, 130 small forwards, 138 power forwards and 153 centers.

Figure 2 - Career Analysis Data Set by Position

By Position	
PG	17.8%
SG	19.4%
SF	19.3%
PF	20.5%
C	22.7%

The following figure outlines the definitions for the variables gathered for the career analysis.

Figure 3 - Career Analysis Variable Descriptions

Variable Descriptions	
Name	**Description**
FG%	Field goal percentage for a player's career.
FT%	Free throw percentage for a player's career.
MPG	Average minutes played per game during a player's career.
PPG	Average points per game during a player's career.
TRB	Average total rebounds per game during a player's career.
AST	Average assists per game during a player's career.
Win Shares	Win shares accumulated over a player's career.
Salary	Dollars accumulated over a player's career.
Pick	Position a player was selected in the draft.

The summary of the statistics gathered for the career analysis is listed below.

Figure 4 - Career Analysis Data Set Summary Statistics

Summary Statistics					
	Max	**Mean**	**Median**	**Min**	**StDev**
FG%	61.60%	44.70%	44.10%	30.30%	4.51%
FT%	91.00%	72.72%	73.90%	44.50%	8.92%
MPG	41.10	21.23	20.70	5.10	7.80
PPG	27.10	8.41	7.40	1.40	4.66
TRB	12.70	3.70	3.20	0.50	2.06
AST	9.90	1.80	1.30	-	1.56
WS	206.4	28.56246291	17.95	-1.6	33.40109394
Salary	$389,250,042.60	$50,758,386.47	$29,714,351.40	$353,592.52	$58,005,977.30
Pick	23.68	22.67	22.31	22.01	21.74

Regression Models

Salary on Win Shares

Results

$$WS = B_0 + B_1 \times Salary \tag{2}$$

By running a regression of salary's effect on win shares, one can see in Figure 6 that 82.40% of the variation in win shares is accounted for by salary. The model as a whole is significant with an F-statistic result of 3,146.15. Additionally, salary is statistically significant with a t-statistic of 56.09 and a p-value of 0. The relationship between salary and win shares indicates that players who are paid more produce more win shares for their respective organizations.

Figure 5 - Salary on Win Shares Results

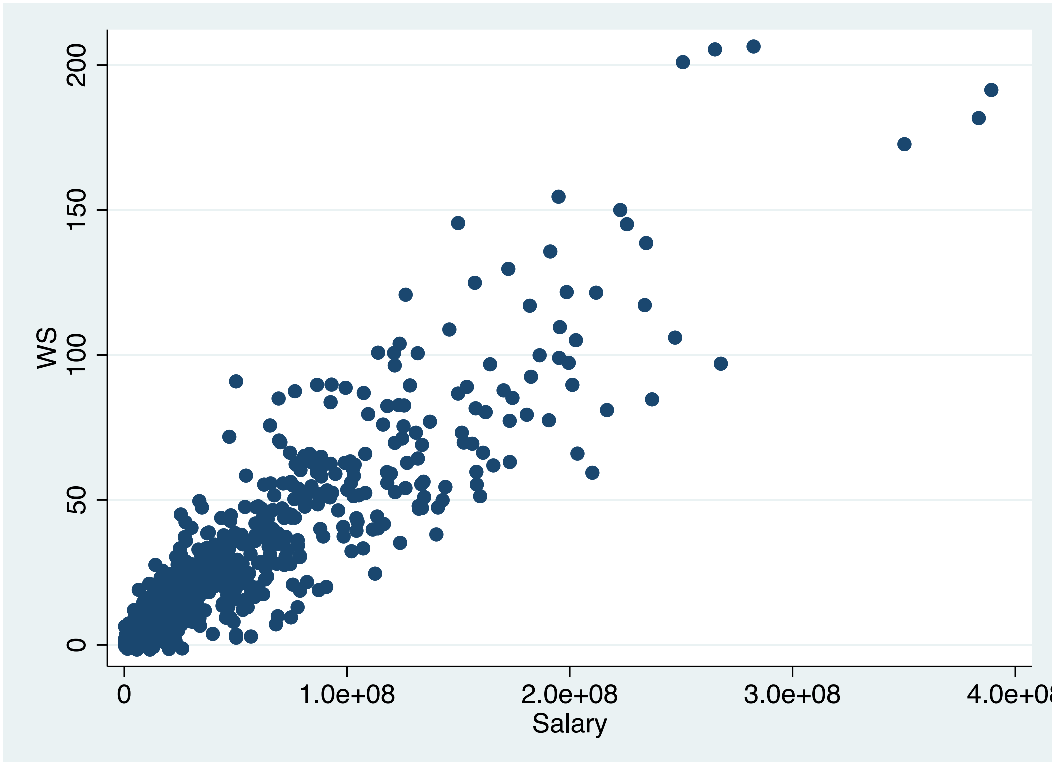
Observations	674
F (1, 672)	3146.15
Adjusted R Squared	82.40%

WS	Coefficient	Std. Error	t	P>(t)	95% Confidence Interval	
Salary	5.23E-07	9.32E-09	56.09	0	5.04E-07	5.41E-07
Constant	2.03	0.72	2.83	0.005	0.62	3.44

Findings

It seems there is a positive linear relationship between salary and win shares. The following figure displays this relationship, however the majority of the observations are located in the bottom left corner of the chart. This depiction shows that there is a talent scarcity issue in the NBA, especially considering this data set includes player career history from eighteen draft classes. The goal of drafting players and making sound team option decisions then becomes more important, so that organizations can avoid capital intensive processes of acquiring players through trades and free agency.

Figure 6 - Salary & Win Share Relationship



Game Play on Win Shares

Results

$$WS = B_0 + B_1 \times FG\% + B_2 \times FT\% + B_3 \times PPG + B_4 \times TRB + B_5 \times AST + B_6 \times MP \quad (3)$$

Game play statistics were gathered for the analysis to determine if it is possible to predict win-shares based on in game performance. This regression shows that the game play statistics measured account for 72.08% of the variation in win shares. All of the variables were statistically significant with large t-statistics and low p-values all below 0.01. Therefore, it is plausible to say that win shares can be predicted based on the game play statistics, which provide insight on in-game performance. This makes sense because the win share statistic is calculated using game play statistics. Although the win share statistic at the base level is a combination of various advanced metrics, those advanced metrics are composed of underlying game play statistics.

Figure 7 - Game Play on Win Shares Results

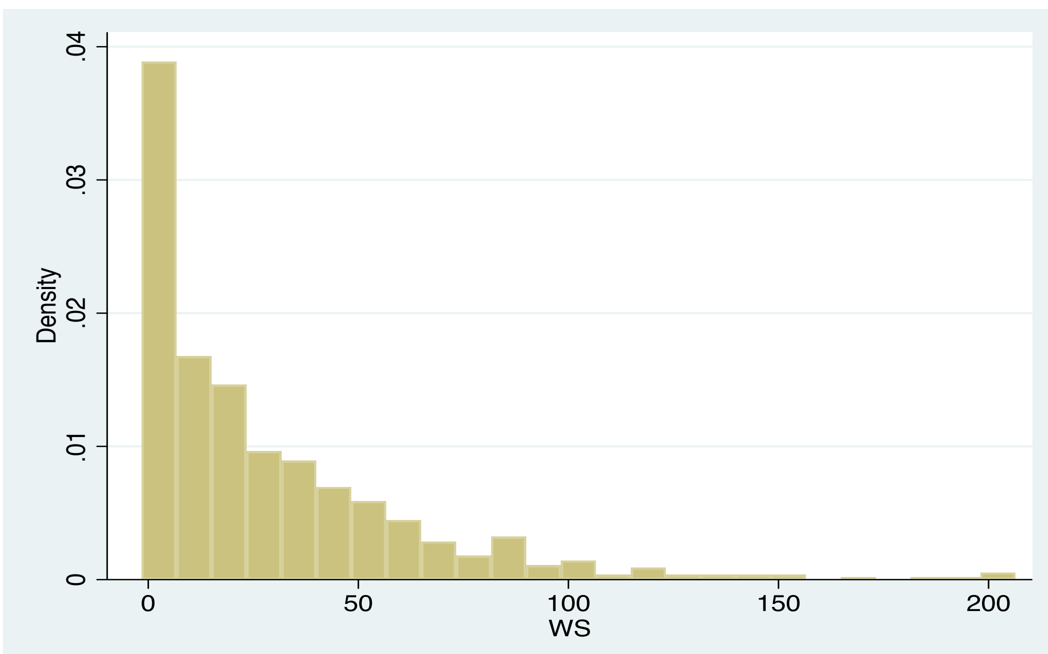
Observations	674
F (6, 667)	286.93
Adjusted R Squared	72.08%

WS	Coefficient	Std. Error	t	P>(t)	95% Confidence Interval	
FG%	52.54	19.42	2.71	0.007	14.41	90.67
FT%	37.33	9.99	3.74	0	17.71	56.95
PPG	3.41	0.38	8.9	0	2.66	4.16
TRB	7.31	0.65	11.26	0	6.04	8.59
AST	7.5	0.73	10.28	0	6.06	8.93
MP	-0.92	0.28	-3.28	0.001	-1.47	-0.37
Constant	-71.78	11.71	-6.13	0	-94.78	-48.78

Career Analysis Discussion

The following figure reinforces the idea that high win share producers are a rare commodity, as the large majority of observations produce close to zero win shares.

Figure 8 - Win Share Distribution



Examining win shares at a career level is important because it allows the completion of two goal-necessary steps outlined in the "Basis of Research" : 1) High win share producers demand higher salaries 2) Win share production levels can be predicted by game play statistics. However, NBA organizations form contracts with players that are time bound. Additionally, said organizations compete to win championships on a year-by-year basis. Therefore, it is important to understand the dynamics of how win share production levels fluctuate over the course of a player's career. To examine this, panel data is necessary.

CHAPTER FIVE

Initial Random Sample Analysis

Description of Purpose

Of the master data set, thirty-five players were randomly selected, representing approximately 5% of the original sample. Seasonal data was collected for these players, creating a panel data set. This panel data set was created to examine the impact of given variables on win shares over the course of a player's career. The master data set essentially gives insight to the question: What do high win share producers look like? While the panel data set attempts to answer: When can teams get a good deal on a high win share producer?

Data Summary

The variables collected for the random sample analysis are described below. One observation is equivalent to one year of data for a given player. Therefore, TRB through PTS are variables that count totals during a given year, as opposed to per-game statistics that are commonly used when analyzing NBA players. The decision to use season totals instead of seasonal per-game averages was largely made to match the data collected for college players.

Figure 9 - Panel Analysis Variable Descriptions

Variable Descriptions	
Name	**Description**
ID	The numerical identifier associated with a player's data
Time	The year of the observation. For example, "1" would indicate a player's rookie year
FG%	Field goal percentage for a given observation
TRB	Total rebounds for a given observation
AST	Total assists for a given observation
STL	Total steals for a given observation
BLK	Total blocks for a given observation
TOV	Total turnovers for a given observation
PF	Total personal fouls for a given observation
PTS	Total points for a given observation
Salary	Salary for a given observation
WS/Season	Win shares for a given observation

The summary of the panel data set is outlined below. As is expected from a random sample, there is wide variation in virtually all of the variables seen through the respective standard deviations.

Figure 10 - Panel Analysis Data Summary

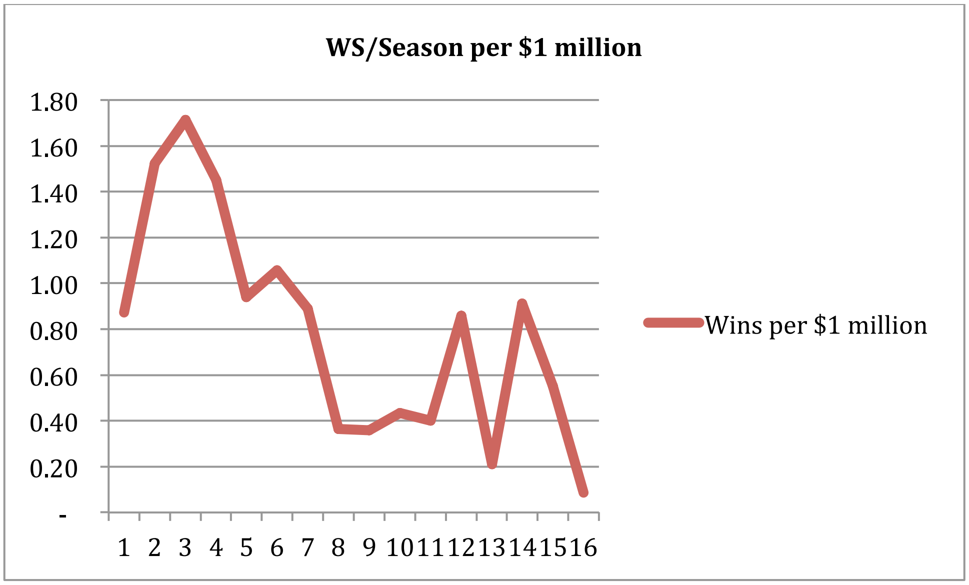
Summary Statistics				
	Mean	**Std. Dev.**	**Min**	**Max**
ID	18.14	10.21	1	35
Time	5.64	3.54	1	16
FG%	0.45	0.06	0.25	1
TRB	256.64	203.53	0	904
AST	96.15	94.91	0	619
STL	43.38	34.64	0	197
BLK	29.68	33.58	0	215
TOV	77.99	59.45	0	315
PF	129.52	76.91	0	371
PTS	548.63	414.63	2	1686
Salary	2.67	2.74	-0.8	12.7
WS/Season	4,495,457	4,478,880	224,018	22,800,000

WS/Season per $1 Million

The figure below shows the average of win shares divided by salary, multiplied by one million, for the entirety of the panel data set. Titled "Wins per $1 million" the chart shows how many wins (Y-Axis) a player will, on average, contribute per $1 million they receive in compensation over time (X-Axis). Essentially, players with high "Wins per $1 million" statistics contribute more wins for lower cost than players with low statistics. Interestingly, the chart peaks at year three. This finding indicates that players contribute most, relative to their compensation, in the third year of their NBA career. Rookie contracts include a team option for the third year, so players that have this option exercised, on average, deliver more value than players not bound by a rookie contract extension. This relationship is seen in the decline after the third year. Teams do have a fourth year option with rookie contracts, so it is interesting not to see the peak in year four.

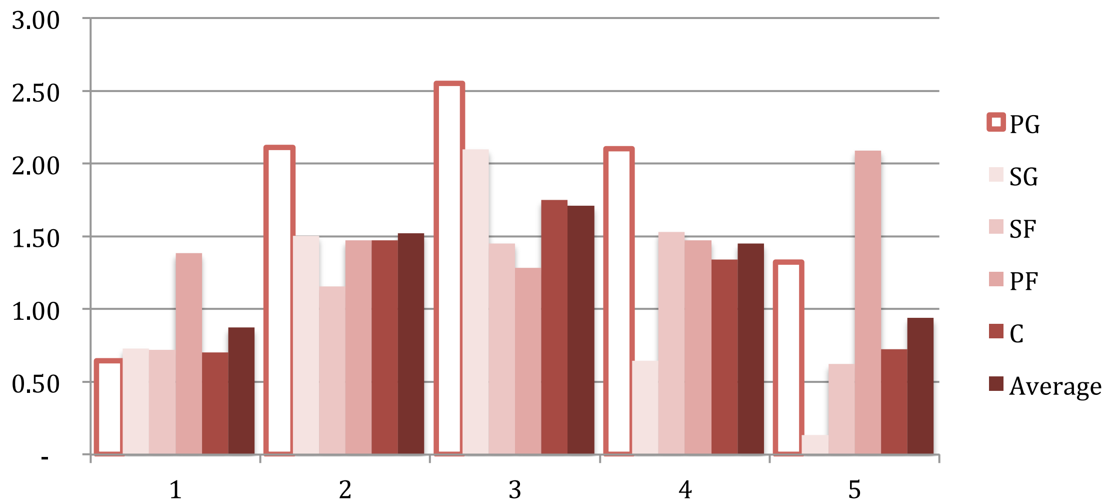
The line roughly bottoms out between years eight and nine. To note, there is possible survivorship bias to explain the spikes in later years, as the players who had longer careers tended to be perennial all-stars.

Figure 11 – WS/Season per $1 million



The following figure displays the information presented above, but by position for the first five years. It is clear that the gap between production and compensation begins to close after year three, since by year five all rookies are either no longer in the NBA or have signed long term contracts. The only position to not follow the trend is power forward, as the average of WS/Season per $1 million for power forwards in the random sample increases in year five. However, the N size for power forwards is only nine and by year five only four players remained in the NBA. Therefore, the high win share production of former perennial all-star Shawn Kemp influences the attribution. In fact, Kemp contributed over seven win shares in year five, while the remaining three power forwards did not even have one win share.

Figure 12 - Random Panel WS/Season per $1M by Position



Initial Random Sample Discussion

The initial analysis of the panel data sought to answer: When can teams get a good deal on a high win share producer? Realizing players contribute the most win shares for the lowest cost in the option years of their rookie contracts, NBA organizations should try hardest to identify high win share producers before they enter the NBA. Acquiring talent early in their career post-draft is difficult due to the low risk NBA teams exert by taking on a player with minimal costs for two years. The following section outlines two analyses that examine how draft position affects win shares over an entire career and how differences in draft position affect win shares at varying career intervals.

CHAPTER SIX

Draft Analysis

Summary

As shown through the "Initial Random Sample Analysis" players are most valuable in terms of win production levels in the team option years (three and four) of their rookie contracts. By selecting players in the draft who are predicted to be high win share producers, teams have the potential of high production for low cost in years three and four. This potential comes at a low risk as teams can discard of underperforming players after year two.

Preliminary Draft Analysis

Description of Purpose

The purpose of this analysis is to determine the relationship between draft position and career win shares and show whether or not players drafted earlier contribute more wins over their respective careers.

Data Summary

The variables collected for this data set are defined below:

Figure 13 - Preliminary Draft Analysis Variable Description

Variable Descriptions	
Name	**Description**
Pick	Draft position in which a player was selected
PickSq	Draft position squared
Avg. Salary	The quotient of career compensation and years played
Career Length	Number of years played in a career
Games	Number of games played in a career
Avg. Games	The quotient of number of games played and number of years played
Minutes	Number of minutes played in a career
Avg. Minutes	The quotient of number of minutes played and number of games played
WinShares	The number of wins attributed to a single player in a career
WinShares/48	The number of wins attributed to a single player per 48 minutes played
College	Dummy variable indicating whether or not a player went to college
Foreign	Dummy variable indicating whether or not a player originated from a foreign country

This data set features data from *Basketball-Reference.com*, including all players drafted from 1989 to 2002. Some observations are missing data for certain variables, such as “Games” and “WinShares”, as some players who are drafted do not play a complete season in the NBA (if at all). Therefore, the regression ran using this data set only had 625 observations, as opposed to the total 728 observations.

Figure 14 - Preliminary Draft Analysis Data Summary

Summary Statistics					
	Observations	**Mean**	**Standard Deviation**	**Minimum**	**Maximum**
Pick	728	29	16	1	58
Avg Salary	728	2970386	3540878	0	2.02E+07
Career Length	625	8	5	1	21
Games	625	427	363	1	1462
Avg Games	728	40	24	0	78.53
Minutes	624	10697	11686	3	50418
Avg Minutes	626	18.32	9.08	0	41.1
WinShares	625	22.49	32.7	-1.6	206.4
College	728	0.91	0.28	0	1
Foreign	728	0.07	0.26	0	1

Regression Model: Pick on Win Shares

Results

$$WS = B_0 + B_1 \times Pick + B_2 \times Pick^2 \quad (4)$$

The figure below highlights the polynomial relationship between “Pick” and “WinShares”, which provided the justification for including the “PickSq” variable.

Figure 15 - Win Shares & Pick Relationship

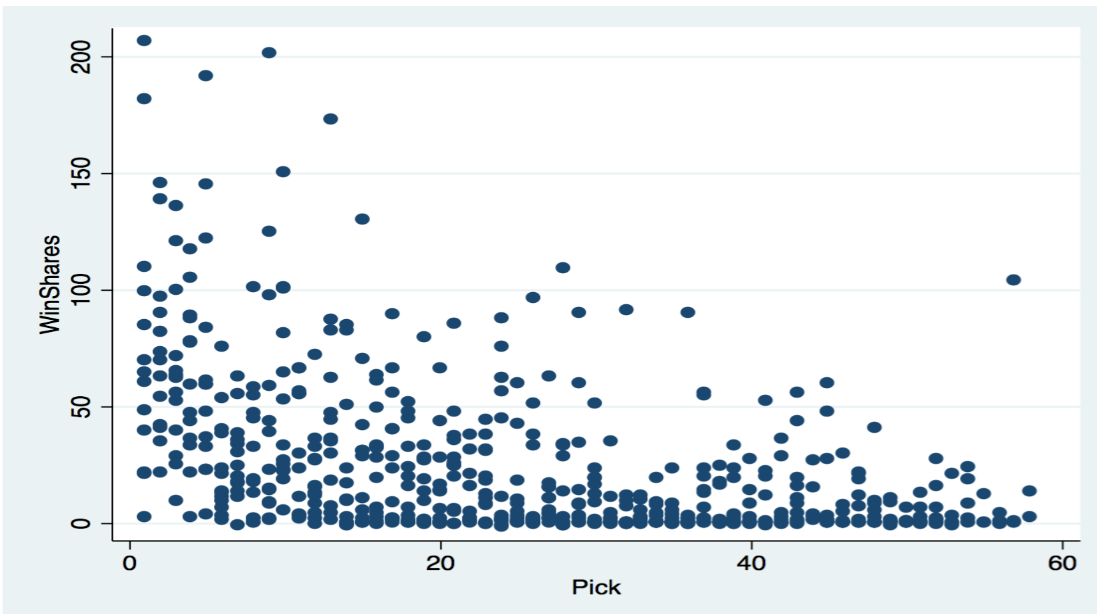


Figure 16 - Pick on Win Share Results

Observations	625
F (2, 622)	126.06
Prob > F	0
Adjusted R Squared	28.61%

WS	Coefficient	Std Error	t	P>|t|	95% Confidence Interval	
Pick	-2.994	0.274	-10.92	0	-3.533	-2.456
PickSq	0.037	0.005	7.52	0	0.0269	0.0461
Constant	66.62	3.25	20.49	0	60.235	73.006

This regression analyzes the relationship between draft position and career win-shares. The model as a whole is significant due to the F-statistic of 126.06. All variables are statistically significant and "Pick" seems to be practically significant because a one-unit increase in draft numerical value results in three less wins contributed by the player. For example, the model suggests a player selected first would contribute almost 64 wins throughout his career, while the player selected last overall would only contribute roughly 17 wins. The R-Squared for this model is 28.84%, indicating that the position a player is selected does not account for the entire variation in win shares. Despite the low R-Squared figure, the model still provides important insight on the relationship.

Findings

The figure below depicts the percentage of players for a specific draft position (over the 14 drafts in which data was collected) who contributed more wins than the regression model predicted. Outperformance was determined by examining the residuals of the regression analysis. Picks one through five outperformed the model at least 50% of the time, while the only other lottery pick to post similar outperformance was pick 13. However, when put into the perspective of the entire draft, picks 21 and 37 were the only picks outside the lottery to outperform the model at least 50% of the time. Therefore, 75% of the draft positions that had a majority of players outperform their predicted win share totals were within the draft lottery, giving credence to the lottery itself.

Figure 17 - Pick on Win Shares Residuals

Percent of Picks with Higher Win Shares vs. Model Prediction							
Pick 1	54%	Pick 15	38%	Pick 30	46%	Pick 45	23%
Pick 2	62%	Pick 16	46%	Pick 31	15%	Pick 46	15%
Pick 3	54%	Pick 17	46%	Pick 32	23%	Pick 47	38%
Pick 4	54%	Pick 18	31%	Pick 33	15%	Pick 48	23%
Pick 5	54%	Pick 19	31%	Pick 34	23%	Pick 49	15%
Pick 6	15%	Pick 20	23%	Pick 35	15%	Pick 50	0%
Pick 7	15%	Pick 21	69%	Pick 36	8%	Pick 51	8%
Pick 8	31%	Pick 22	23%	Pick 37	54%	Pick 52	15%
Pick 9	38%	Pick 23	46%	Pick 38	23%	Pick 53	8%
Pick 10	46%	Pick 24	38%	Pick 39	23%	Pick 54	15%
Pick 11	38%	Pick 25	23%	Pick 40	23%	Pick 55	0%
Pick 12	8%	Pick 26	31%	Pick 41	31%	Pick 56	0%
Pick 13	62%	Pick 27	31%	Pick 42	15%	Pick 57	14%
Pick 14	23%	Pick 28	38%	Pick 43	46%	Pick 58	0%
		Pick 29	31%	Pick 44	15%		

Draft Position Analysis

Description of Purpose

Knowing that players are paid least for production in the early years of their career, and that teams subsequently must draft well, this analysis seeks to examine whether a relationship exists between where a player is drafted and the win shares they produce over the course of their career. The dataset for this analysis is a subset of the dataset collected for the “Career Analysis”, using all first, fifteenth, thirtieth, forty-fifth and sixtieth draft selections made over the course of the eighteen drafts from 1989 to 2006. However, this dataset was converted from career data to season-by-season data for win share and salary statistics to create a panel data set.

Data Summary

The descriptions of the variables are listed below. Important to note is the "Last" variable, which signifies players drafted last in their respective draft. The last pick number varied throughout the data collected from 1989 to 2006 due to NBA expansion.

Figure 18 - Draft Position Variable Descriptions

Variable Descriptions	
Name	**Description**
Time	Describes the season of a player's career (1 represents a rookie season)
WS/Season	Describes the win shares produced per season
Salary/Season	Describes compensation per season
First	Describes players drafted first overall
Fifteenth	Describes players drafted in the fifteenth position
Thirtieth	Describes players drafted in the thirtieth position
Forty-Fifth	Describes players drafted in the forty-fifth position
Last	Describes players drafted last overall

The summary of the data is shown below. There seems to be great variation in win shares among observations. Also, as a consequence of later draft picks being more likely to not have an NBA career (or shorter careers in general), the dataset was skewed toward earlier picks in terms of observations.

Figure 19 - Draft Position Data Summary

Summary Statistics					
	Observations	**Mean**	**Standard Deviation**	**Minimum**	**Maximum**
Time	625	6.48	4.15	1	19
Pick	625	18.56	17.94	1	60
WS/Season	625	3.95	4.03	-1.2	20.3
Salary/Season	625	7,112,345	7,070,432	28,751.5	40,300,000
First	625	0.39	0.49	0	1
Fiftienth	625	0.23	0.42	0	1
Thirtieth	625	0.19	0.39	0	1
Forty-Fifth	625	0.15	0.36	0	1
Last	625	0.04	0.20	0	1

Regression Model: Draft Position on WS/Season

A panel regression was run for this analysis, setting Player ID (a numerical value assigned to each player for identification, not analysis, purposes) and Time for the panel. A fixed effect was not used due to dummy variables being created manually for the pick groups. The first overall picks were left out of the regression to be used as the comparison group.

Results

$$\frac{WS}{Season} = B_0 + B_1 \times Fifteenth + B_2 \times Thirtieth + B_3 \times Forty - Fifth + B_4 \times Last \tag{5}$$

The figure below displays the results from the draft position regression. While the r-squared value is low, with only 16.33% of the variation in win shares explained by the draft groups, it is important to note that all of the variables were extremely statistically and practically significant. Interestingly, the coefficients on the "Fifteenth", "Thirtieth and "Forty-Fifth" variables are all relatively similar. Therefore a player drafted 15^{th}, 30^{th} or 45^{th} in the draft should perform similarly to one another compared to players drafted first overall, ceteris paribus. The average for WS/Season in the dataset was 3.95. Using the regression model, it is impossible to arrive at a WS/Season statistic equal to, or above, the average. With the mean win shares for first overall picks at 5.92, it is clear that the comparison group skewed the overall dataset.

Figure 20 - Draft Position Regression Results

Observations	625
Groups	79
Adjusted R Squared	16.33%

WS/Season	Coefficient	Std Error	z	P>|z|	95% Confidence Interval	
Fifteenth	-3.40	0.69	-4.91	0.00	-4.76	-2.04
Thirtieth	-3.62	0.72	-5.04	0.00	-5.03	-2.21
Forty-Fifth	-3.63	0.74	-4.91	0.00	-5.08	-2.18
Last	-4.87	1.06	-4.60	0.00	-6.94	-2.80
Constant	5.63	0.47	12.05	0.00	4.71	6.54

Findings

The regression may not have shown that draft position is a large determinant of win shares, however it does lend credence to the findings of the "Preliminary Draft Analysis". That model showed that outside the first few picks, there are relatively few positions that outperform predicted win share levels historically. This model shows that there is relative parity among all of the draft groups except the "Last" group when comparing performance to first overall picks. Essentially, this means that high win share producers are most likely found at the top of the draft, but can be found throughout the draft and picks outside the lottery are all roughly similar in value. Do NBA teams currently act accordingly to the results of this regression? It is hard to say with certainty, as historical draft moves are not formally documented. However, in the 2017 draft there were six moves made to either buy into or shift position in the second round (*NBA.com*, 2017). This activity runs contrast to the data provided above, as picks outside the first fourteen are relatively equal.

Draft Discussion

The "Initial Random Sample Analysis" provided evidence that players produce more win shares per million dollars during the early years of their career. The way teams acquire rookies is through the draft. Hence, a logical follow up question is: Where are high win share producers historically selected in the draft? The "Preliminary Draft Analysis" displayed that, outside the first five selections, players drafted at nearly all other draft positions did not outperform the model in terms of win shares historically. Therefore, the model seems to show that the first five draft picks are positions where high win share producers have been selected historically.

Additionally, the "Draft Position Analysis" provided evidence that first overall picks, when compared to the remaining draft groups, do produce more win shares. However, the middle three draft groups (15^{th}, 30^{th} and 45^{th}) perform similarly when compared to first overall picks, meaning that if teams do not secure a lottery pick, it is not essential to "move up" in the draft and forfeit assets like veterans or future draft picks. While teams are greatly benefited by superstar players that contribute outlier win share statistics, it is also necessary for teams to round out rosters with players who routinely provide positive win shares.

CHAPTER SEVEN

Secondary Random Sample Analysis

Description of Purpose

The "Draft Analysis" showed that high win share producers are historically selected earlier in the draft, however teams do not perennially have a top five or even top fifteen pick. Therefore, it is important to understand how to make sound draft selections with later picks. To accomplish the stated task, it is important to understand how to predict player win production levels using only data available during a player's collegiate career. Additionally, once a player is drafted, determining whether to exercise team options in years three and four is essential to the talent management process of NBA organizations. Using the random sample described in the "Initial Random Sample Analysis" three regressions were executed to accomplish the aforementioned goals.

Regression Models

Panel Data Regression

The random sample was used in its entirety for the panel regression. This regression was run to show whether or not in-game statistics are sound predictors of win shares per season when analyzed on a season-by-season basis. The variable descriptions and data summary for this regression can be found within the "Initial Random Sample Analysis" section of this paper. A fixed effects panel regression was run for this analysis, setting player id and time for the panel.

Results

$$\frac{WS}{Season} = B_0 + B_1 \times FG\% + B_2 \times TRB + B_3 \times AST + B_4 \times STL + B_5 \times BLK + B_6 \times TOV + B_7 \times PF + B_8 \times PTS \quad (6)$$

The variables selected account for 73.76% of win shares per season. All of the variables are statistically significant with large t values and P values of zero or near zero. The most practically significant variable is FG% because a one percentage point increase in FG% yields 5.404 more win shares per season holding all else constant. This finding lends to the idea that efficient players have higher win shares, due to the minimization of activities that are detrimental to their teams, such as missing attempted shots.

Figure 21 – Panel Regression Results

Observations	292
Groups	35
Adjusted R Squared	73.76%

WS/Season	Coefficient	Std. Error	t	P>(t)	95% Confidence Interval	
FG%	5.006	1.261	3.97	0.000	2.5218	7.4900
TRB	0.003	0.001	2.63	0.009	0.0007	0.0049
AST	0.005	0.002	2.61	0.010	0.0013	0.0096
STL	0.012	0.006	1.94	0.053	-0.0002	0.0240
BLK	0.012	0.004	2.77	0.006	0.0034	0.0204
TOV	-0.016	0.005	-3.50	0.001	-0.0255	-0.0071
PF	-0.008	0.002	-3.12	0.002	-0.0125	-0.0028
PTS	0.005	0.001	7.15	0.000	0.0036	0.0064
Constant	-2.163	0.549	-3.94	0.000	-3.2436	-1.0825

Findings

This model is a better representation of the relationship between in-game statistics and win shares than "Game Play on Win Shares" displayed in the career analysis, although it reaffirms the findings. Intuitively, because the win share statistic is built off of in-game statistics, it makes sense that in-game statistics would be good predictors of win production levels.

College Data Regression

For the college data regression, only the last year of college data prior to a player entering the NBA draft was used. Additionally, the win shares per season for the rookie year observations were used as the Y variable, to determine whether or not college in-game statistics can determine rookie production levels. This regression is significant because NBA organizations must draft well to maximize production compared to cost of roster talent as seen in the "WS/Season per $1 Million" analysis.

Data Summary

Of the thirty-five players included in the panel regression, only thirty-three players were used in this model, as two of the observations did not attend college. Since this model includes two fewer observations, an additional data summary is provided below for accuracy. All variable descriptions remain the same from those assigned in the "Initial Random Sample Analysis". "TOV" and "PF" were not included in the model for college data as many players were missing this information, indicating that turnovers and personal fouls are statistics that have not been tracked in college basketball until recently.

Figure 22 - College Data Summary

Summary Statistics				
	Mean	**Std. Dev.**	**Min**	**Max**
FG	0.4936	0.0493	0.42	0.65
TRB	219.85	73.57	78	352
AST	82.18	64.11	9	299
STL	44.06	27.08	5	111
BLK	33.88	34.27	0	156
PTS	567.82	145.52	135	818
WS/Season Rookie	0.88	1.22	-0.50	4.60

Results

$$\frac{WS}{RookieSeason} = B_0 + B_1 \times FG + B_2 \times TRB + B_3 \times AST + B_4 \times STL + B_5 \times BLK + B_6 \times PTS \quad (7)$$

The regression results below indicate that college in-game statistics do not account for any of the variation in rookie win shares when viewing adjusted r-squared. None of the variables were statistically or practically significant.

Figure 23 - College Data Regression Results

Observations	33
F (9, 23)	0.34
Adj. R-Squared	-14.22%

WS/Season Rookie	**Coefficient**	**Std. Err.**	**t**	**P>(t)**	**95% Confidence Interval**	
FG	0.609	5.524	0.11	0.913	-10.75	11.96
TRB	-0.003	0.005	-0.70	0.488	-0.01	0.01
AST	0.003	0.006	0.49	0.630	-0.01	0.02
STL	-0.017	0.015	-1.12	0.273	-0.05	0.01
BLK	0.005	0.009	0.60	0.553	-0.01	0.02
PTS	0.001	0.002	0.41	0.687	0.00	0.01
Constant	1.150	2.772	0.4	0.682	-4.55	6.85

Findings

Although the regression did not yield significant results, the model shows that in-game statistics for the last year of collegiate play are not a sound indicator of rookie win share totals. Examining more than one year of collegiate play could be beneficial, however a large share of drafted players only play one year of college basketball. This model does not factor in the conference a college player was in, nor their team's strength of schedule, as Alexander Greene did in his model. These two data points are important, as there is disparity among conferences and schedules in college basketball. While Greene's model aimed to look at different relationships, including more qualitative data for the college observations may improve this model. Understanding that drafting well is important in taking advantage of the most productive years of a player's career in terms of wins per million dollars, it seems NBA organizations must find other ways of determining rookie win share totals from data available for collegiate athletes.

Despite the shortcomings of this model, examining whether third year win shares can be determined by data available for rookies is still mission critical. The following model seeks to answer this, moving closer toward the overarching goal of the paper, which is to determine whether NBA organizations can make better decisions regarding rookie contracts in option years by analyzing win shares.

Rookie Data Regression

For the rookie data regression, only the year one observations for each player in the random sample were used. Additionally, the win shares for third year observations were used as the Y variable, to determine whether or not rookie in-game statistics can determine third year production levels. This regression is significant because NBA organizations must make decisions regarding third year team options for rookies before the start of the second season. Therefore, teams are making these decisions based solely on rookie data.

Data Summary

Of the thirty-five players selected for the random sample, only thirty-three played three or more seasons and therefore only thirty-three observations are found in this regression. Since the observations for this regression were altered from the original model, a new data summary was generated for accuracy. The variable names all remain valid from the "Initial Random Sample Analysis".

Figure 24 - Rookie Data Summary

Summary Statistics				
	Mean	**Std. Dev.**	**Min**	**Max**
FG	0.4361	0.0597	0.29	0.57
TRB	160.94	135.48	4	391
AST	50.58	52.21	3	205
STL	25.45	20.69	0	78
BLK	24.94	38.09	0	163
TOV	56.45	46.31	2	163
PF	98.97	76.96	0	230
PTS	330.91	292.01	23	1142
WS/Season Rookie	0.911	1.237	-0.50	4.60
WS/Season Year 3	2.855	2.874	-0.30	11.30
Salary (Rookie)	1,837,976	1,028,515	507,239.30	4,759,380

Results

$$\frac{WS}{Year3Season} == B_0 + B_1 \times FG + B_2 \times TRB + B_3 \times AST + B_4 \times STL + B_5 \times BLK + B_6 \times PTS + B_7 \times \frac{WS}{RookieSeason} + B_8 \times RookieSalary \quad (8)$$

As seen in the figure below, the variables used in the regression account for 45.20% of the variation in win shares per season. Therefore, rookie data only accounts for roughly 30% of win shares produced in a player's third season. The R-Squared for this regression was 62.33% so it seems the regression is being penalized for including too many variables.

"FG" seems to be practically significant but not statistically significant, as the t-statistic and p-value are not significant, but an increase of one percentage point in field goal percentage in a player's rookie season yields 6.92 more win shares in a player's third season. Interestingly, "PF" seems to be statistically significant, alluding to the fact that players who commit less personal fouls in their rookie season produce more win shares in their third season. In fact, the variable seems to hold practical significance, as the mean in the data summary is 98 personal fouls. The average personal foul total would yield 3.64 win shares less in a player's third season, which is greater than the standard deviation for win shares. "PTS" is statistically significant but not practically significant, as the coefficient is negative. Since the team with more points wins, it is unrealistic that scoring more points in a rookie season would result in less win shares in a third season. A possible explanation to this result is that players who excel in their rookie season are defended with more urgency in subsequent seasons. "WS/Season Rookie" seems to be statistically significant with a t-statistic above two and a p value near zero. The variable also seems to be practically significant as one additional win share in a player's rookie season produces roughly 1.5 more win shares in a player's third season holding all else constant.

Figure 25 - Rookie Data Regression Results

Observations	33
F (9, 23)	3.64
Adj. R-Squared	45.20%

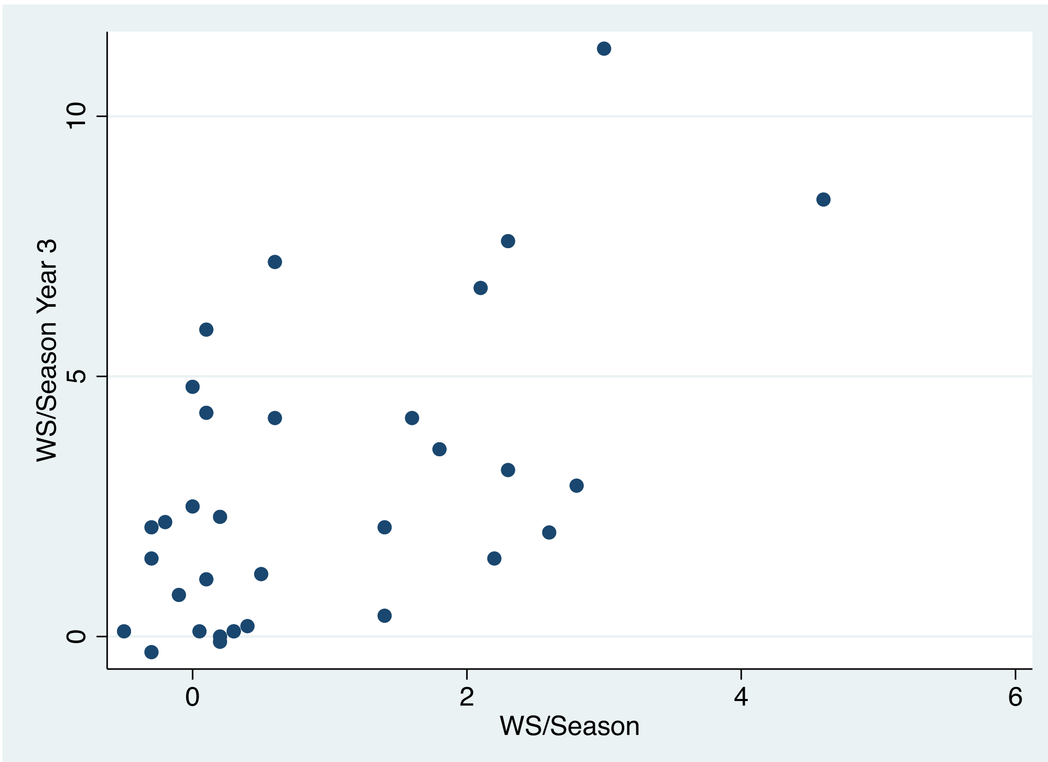
WS/Season Year 3	**Coefficient**	**Std. Err.**	**t**	**P>(t)**	**95% Confidence Interval**	
FG	6.92477	7.881	0.88	0.389	-9.420	23.269
TRB	0.00057	0.012	0.05	0,964	-0.026	0.027
AST	-0.00062	0.022	-0.03	0.978	-0.047	0.046
STL	0.03929	0.057	0.68	0.501	-0.080	0.159
BLK	0.00773	0.017	0.46	0.652	-0.027	0.043
TOV	0.09750	0.053	1.82	0.082	-0.014	0.209
PF	-0.03710	0.014	-2.58	0.017	-0.067	-0.007
PTS	-0.01226	0.006	-2.00	0.058	-0.025	0.000
WS/Season Rookie	1.48867	0.549	2.71	0.013	0.350	2.628
Salary (Rookie)	1.06E-06	5.59E-07	1.90	0.071	-9.79E-08	2.22E-06
Constant	-2.07071	3.427	-0.60	0.552	-9.179	5.037

Findings

The regression may not account for all the variation in year three win shares, but it does yield interesting results. Needing an F statistic of 2.32 to be significant, the result of 3.64 shows that the overall model is statistically significant. "WS/Season Rookie" was both statistically and practically significant showing that win shares produced in a player's rookie season are a good indicator of year three win shares. Additionally, "PF" was both statistically and practically significant and was negative, alluding to an inverse relationship between rookie personal fouls and year three win shares. This relationship makes sense and its significance shows that teams should observe how controlled a player is on the court. However, this regression is not a perfect model. The model does not account for all of the variation in year three. One reason could be win shares are the learning curve between college and the professional level. Players who do not perform well in their rookie season could excel by year three. Those who outperform in their rookie season may see opponents adjust and face increased competition by their third year.

The figure below shows the relationship between rookies' win-shares and win-shares in year three.

Figure 26 - WS/Season Rookie & WS/Season Year 3 Relationship



Secondary Random Sample Discussion

It is evident through the results of the panel regression that in-game statistics correlate with win shares over the course of a player's career. However, using the same statistics from college do not seem to show the same relationship with win shares for players during their rookie season. Interestingly, using the same in-game statistics with the addition of rookie win shares and rookie salary seem to give some indication of year three win shares. The last point is of most importance to this paper, as the relationship between in-game statistics and year three win shares show that NBA organizations can make better judgments about the team options in rookie contracts by examining similar data.

By incorporating win shares per one million dollars analysis, teams can then determine whether a player is contributing production that exceeds their cost. While teams do not have much leeway in what a player's third year salary is, due to rookie contract constraints, they do have flexibility in the qualifying offer extended after a player's fourth year. Since the qualifying offer must be made before the beginning of the fourth year, teams must decide what the player is worth in terms of compensation after year three. Therefore, understanding the projection of a player's development in terms of wins contributed through their first three years is vital in determining the dollar value to assign to a qualifying offer at the end of a player's third year.

CHAPTER EIGHT

Conclusion & Further Research

Conclusion

This paper finds that efficiency variables such as field goal percentage and proxy variables for self-control, like personal fouls, serve as sound predictors of future win share production. Through analysis of win shares versus compensation, it is found that players are typically paid in accordance to their production level, except when constrained by a rookie contract. Therefore, drafting seems to be a better solution than free agency or trades in locating high win share producers. Unfortunately, in-game statistics that serve as relatively good predictors of win share levels in the NBA are not good predictors when collegiate data is used. However, analysis of the draft shows that high win share producers are typically found in earlier selections.

Further Research

Further research could focus on finding a method to predict win shares for players entering the draft using data available on the collegiate level. Succeeding on this front would allow for perfect information in the draft market and eliminate errors like the Michael Jordan incident. NBA organizations are ultimately profit seeking firms and minimizing risks that result in sunk costs is highly beneficial. Additionally, further research could seek to examine adequate compensation levels for qualifying offers based on data available in the first three years of a player's career. This analysis would show how NBA teams can take the team option analysis a step further and not overpay to keep talent, effectively eliminating the drop off in win shares per million dollars. Finally, from a different perspective, further research could take analysis from this paper relating to the disparity in production versus compensation under rookie contracts, along with the research in the "Joshi Analysis" to determine rational changes for the next collective bargaining agreement. It certainly is not in the best interest of players for this gap in production and compensation to exist, and therefore seems like an area for other researchers to examine. However, as long as this gap persists, NBA franchises should seek to exploit it to maximize seasonal wins and, in turn, attempts at championships as post season success seems to be the leading determinant in short and long term economic growth for organizations.

WORKS CITED

"B-R Bullpen." *Baseball-Reference*, www.baseball-reference.com/bullpen/Bill_James.
Cleveland Cavaliers franchise value 2003-2017 Statistic. (n.d.). Retrieved October 21, 2017, from https://www.statista.com/statistics/194649/franchise-value-of-the-cleveland-cavaliers/

Coates, Dennis, and Babatunde Oguntimein. "The Length and Success of NBA Careers: Does College Success Predict Professional Outcomes." *North American Association of Sports Economists* , Aug. 2008, citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.830.7628&rep=rep1&type=pdf.

Dengate, Jeff. "Let the Ping-Pong Balls Fall." *NBA.com*. N.p., 24 May 2005. Web. 27 Apr. 2017.

"Glossary." *Basketball-Reference.com*, www.basketball-reference.com/about/glossary.html.

Greene, Alexander C. "The Success of NBA Draft Picks: Can College Careers Predict NBA Winners." *TheRepository at St. Cloud State*, Nov. 2015, repository.stcloudstate.edu/cgi/viewcontent.cgi?article=1002&context=stat_etds.

"Hakeem Olajuwon." *Biography.com*, A&E Networks Television, 4 Feb. 2016, www.biography.com/people/hakeem-olajuwon-21101363.

James, B., & Henzler, J. (2002). Win Shares. STATS LLC.

Hausman and Leonard (1997), Superstars in the National Basketball Association: Economic Value and Policy, *Journal of Labor Economics* 15, no. 4: 586-624.

Jessop, Alicia. "The Structure of NBA Rookie Contracts." *Forbes*, Forbes Magazine, 28 June 2012, www.forbes.com/sites/aliciajessop/2012/06/28/the-structure-of-nba-rookie-contracts/#31d1226c7299.

Joshi, Nikhil. "The Market for Talent: Are Top Draft Picks Overpaid." *Stanford University Department of Economics*, May 2011, economics.stanford.edu/sites/default/files/publications/nikhil_hthesis2011.pdf.

Kubatko, Justin. "Calculating Individual Offensive and Defensive Ratings." *Basketball-Reference.com*, www.basketball-reference.com/about/ratings.html.
Massey, C., & Thaler, R. (2005). The Loser's Curse: Decision Making and Market Efficiency in the National Football League Draft. Retrieved October 21, 2017, from https://www.chicagobooth.edu/alumni/downloads/thaler_losers_curse.pdf.

"NBA Draft Rules." *DraftSite: The Original Full Round Mock Draft Site*. N.p., n.d. Web. 27 Apr. 2017.

NBA Win Shares. (n.d.). Retrieved October 21, 2017, from https://www.basketball-reference.com/about/ws.html

Nusser, Jeff. "Recruiting: What do all those 'stars' mean?" *CougCenter*, CougCenter, 5 Feb. 2013, www.cougcenter.com/wsu-football-recruiting/2013/2/5/3956800/rivals-scout-espn-247-star-rating-system-national-signing-day.

Oliver, D. (2004). Basketball on Paper. Potomac Books Inc.

Schoenfield, David. "The 100 Worst Draft Picks Ever." *ESPN*, ESPN Internet Ventures, 26 Apr. 1996, www.espn.com/espn/page2/story?page=schoenfield%2F060427.

Silver, Nate, and Neil Paine. "We're Predicting The Career Of Every NBA Player. Here's How." *FiveThirtyEight*, FiveThirtyEight, 23 Oct. 2015, fivethirtyeight.com/features/how-were-predicting-nba-player-career/.

Staff, NBA.com, and NBA.com. "2017 NBA Draft Trade Tracker." *NBA.com*, NBA.com, 21 June 2017, www.nba.com/2017-nba-draft-trade-tracker.

The Business Of Basketball. (2017). Retrieved October 21, 2017, from https://www.forbes.com/nba-valuations/list/#tab:overall

Titus, Mark. "The Problem With the Draymond Prototype." *The Ringer*, The Ringer, 17 May 2017, www.theringer.com/2017/5/17/16046654/draymond-green-golden-state-warriors-nba-draft-3b2fbbcedf76.

Wasserman, Jonathan. "What Happens at the NBA Combine?" *Bleacher Report*. Bleacher Report, 25 Apr. 2017. Web. 27 Apr. 2017.