11-2015

The Success of NBA Draft Picks: Can College Careers Predict NBA Winners?

Alexander C. Greene
St. Cloud State University

The Success of NBA Draft Picks: Can College Careers Predict NBA Winners?

by

Alexander Carl Greene

A Thesis

Submitted to the Graduate Faculty of

St. Cloud State University

in Partial Fulfillments of the Requirements

for the Degree of

Master of Science in

Applied Statistics

November, 2015

Thesis Committee:
David Robinson, Chairperson
William Branson
Michael Ernst

Abstract

This is a study of predicting success in the NBA based on college experience, and re-predicting after the rookie year of professional basketball. All first and second round picks from the 1985 draft through the 2005 draft are included in overall analysis, with 841 players having had at least one year of NBA experience and have played on a Division I NCAA team. The primary endpoints used in analyses are Player Efficiency Rating (PER), win shares, and win shares per 48 minutes. This paper will predict the success of picks using their draft pick, college statistics (both qualitative and quantitative), and physical qualities (height and weight). Also, rookie year statistics are used to update the analysis to determine if any additional information is gained after one year of professional basketball. This study concludes that a statistical analysis of college statistics predicted performance well using win shares per 48 minutes, but did not improve predicted performance with PER and win shares for first round draft picks. In addition, one of the predictive formulas was able to predict performance of the top 100 NBA prospects with higher accuracy than the actual draft.

Acknowledgements

I would like to thank everyone involved in helping me complete this paper. To Dr. Robinson, who helped me with revising the paper many times. To the entire Statistics department at SCSU, who have given me so much help and guidance. Thanks to my beautiful wife and children, for encouraging me to finish my degree. Thanks also to my parents, for always being supportive of me. And finally, thanks to Jesus Christ, my savior.

Table of Contents

List of Tables

Table Page

List of Figures

Chapter 1: Introduction

In the National Basketball Association (NBA), millions of dollars and thousands of man hours are spent researching players eligible to be drafted in the NBA draft. During the draft, teams have the opportunity to choose players to add to their rosters. One of the benefits of the draft is that parity is instilled around the league. Teams with the worst records have a chance to sign the most talented players available in a given year. The teams with the best records have lower draft picks, and thus, usually less talented players to sign. Otherwise, if players were able to sign with whichever teams they wanted, the teams that experienced the most success would be the most popular destinations. With the draft, there is hope for every team to sign a potential superstar.

If the system was perfect, the best player would be taken first, the second best player would be taken second, and so forth. But, as history has shown, there have been many misfires in predicting the talent of NBA prospects. The most notorious pick in the NBA draft belongs to the Portland Trailblazers, who, with the second pick in the 1984 NBA draft, chose Sam Bowie, a center out of Kentucky, over Michael Jordan, who is widely considered to be the best player in the history of the NBA (Golliver, 2012). The Chicago Bulls went on to win six NBA Championships with Jordan, and Jordan retired with five MVP awards, 10 scoring titles, and the highest points per game average in NBA history. Bowie had a lengthy career, but dealt with multiple leg injuries which ultimately diminished his ability to live up to his lofty potential.

Looking at a more recent example, Greg Oden, a center from Ohio State, and Kevin Durant, a forward out of the University of Texas, were widely considered to be the top two prospects in the 2007 NBA draft. Once again, the Portland Trailblazers, who owned the number one overall pick, had to make a difficult decision. Once again, they chose the dominant seven footer over the dynamic scorer (2007 NBA Draft, n.d.). Oden would suffer multiple knee injuries throughout his career, ultimately playing in 105 games. Durant would go on to win four scoring titles and an MVP trophy before his 26th birthday.

There are other instances where teams have passed up future All-Stars to draft players that would ultimately be out of the league in a few years, which leads to the question: Can the way that talent is evaluated leading up to the draft be improved? This paper will put forth analytical tools to better predict which players will become productive NBA players, using Player Efficiency Rating (PER), win shares, and win shares per minute as our criteria. These variables will be defined in the next section of this paper.

One important question to ask at the beginning of the study is who wants to know about this study and what do they want to know about it? Inherently, the people who would benefit from this information would be owners and general managers (GM’s) of the 30 NBA franchises, the people who draft and sign players to multi-million dollar contracts. NBA teams want to win championships. On the court, that is the ultimate goal. This paper will attempt to quantify the most important factors in predicting success in the NBA, using college basketball statistics. It will also predict NBA success after 1 year of experience in the NBA, to see what knowledge, if any, is gained about the success of a player’s career. Ultimately, this paper will put forth additional resources to be used by NBA teams in preparing for the NBA draft. Predicting PER and win shares will be the focus of our analysis.

Chapter 2: Literature Review

An area that has had extensive research in predicting future success is major league baseball (MLB). Since the history of MLB is quite long, there have been many opportunities for people to analyze the huge sets of data that the games produce. Bill James, widely considered the father of modern baseball analytics, began this trend of looking deeper into the data in 1977 with his book *The Bill James Baseball Abstract.* Instead of recapping the games and interviewing the players, James looked at the box scores, and tried to discern patterns hidden in the data. Within a few years, many in the baseball community were taking notice. James introduced new terms, such as Runs Produced and Win Shares, which are commonly used today when evaluating players (Zminda, 2010).

Naturally, since this paper is looking at NCAA basketball statistics to predict NBA success, it makes sense to compare to the work being done in the MLB. The main difference, however, is that MLB actually has a minor league system. There are three levels of minor league baseball: A, AA, and AAA, with A being the lowest and AAA being the highest, one level below MLB. Since the NBA does not have a minor league system, NCAA basketball is the closest thing relatively to minor league baseball.

Just as PER is a catch-all statistic in the NBA, Wins Above Replacement, or WAR, has become the all-in-one statistic used to evaluate MLB players. WAR can be used for both pitchers and position players to determine how many wins a player contributes above what a "replacement", or average, player would provide. With WAR as the primary endpoint for career success, Tymkovich (2012) looks at the pre-draft rankings of players, and compares their career WAR to determine how successfully the players were ranked before they were drafted. It was found that the three top publications for ranking MLB prospects [Baseball America, Baseball Prospectus, and John Sickels (ESPN)] were all successful in predicting players who would have higher career WAR.

There are different ways to measure success in MLB. Chandler and Stevens (2012) created a variable called ML Contribution, which was simply a binary variable with 1 = player made it to MLB, and 0 = player did not make it to MLB. Looking at 1,019 minor league players who played at least 15 minor league games between 1999 and 2002, they used traditional baseball metrics (runs, runs batted in, batting average, etc.) to predict if a player would make it up to the major leagues. Using classification trees, which splits up the sample using different criteria (e.g., AVG > .250 and AVG <= .250), they were able to determine which variables were significant in predicting players moving up to MLB. While none were significant at A level minor league baseball, at AA and AAA, batting average, on-base percentage, slugging percentage, and on-base percentage plus slugging were all significant positive predictors in a player moving up to MLB. This shows not only that players are ranked well before the draft, but that teams can look at concrete numbers from the minor leagues to predict success.

As analysts get smarter about looking deeper into the data for predicting future success in sports, it would naturally follow that certain variables would lose relevancy over time. Chang and Zenilman (2013) decided to split the recent history of baseball into three periods: pre-Moneyball, post-Moneyball, and post-post-Moneyball. Moneyball is in reference to the Oakland Athletics in the 2002 season. Oakland’s management decided to use new analytical methods in their pursuit of new players to acquire. Despite early ridicule, the Athletics went on to have a very successful season, including a 20 game winning streak.

Looking at the pre-Moneyball era, length of a player's contract, a player's height, stolen bases, on-base percentage plus slugging, and how many double plays a player grounded into were all significant variables in predicting salary, which is an indication of the faith an organization has in a player. Once into the post-Moneyball era, height and stolen bases lost their significance. This was due mainly to height being attributed to being better at baseball, and total stolen bases being considered a measurement of speed, regardless of how efficiently those bases were being stolen. Once there had been significant time after Moneyball, the post-post-Moneyball era, in 2011, a completely different picture comes into focus. Only WAR, Length of Contract, and how many double plays a player grounded into were significant in predicting salary. This shows the shift in thinking from using the eye test on a player to really looking at the numbers they are producing on the field. As the sophistication of analysis increases for a sport, the important factors to look at will also change. The work in baseball analytics leads into the work that has been done in the realm of predicting NBA success.

Several studies have undertaken the task of studying the NBA draft for economic purposes, treating the NBA as an enterprise and the players as the employees. Staw and Hoang (1995) were the first to examine sunk costs in the NBA draft. In economics, sunk costs refer to the money already invested in a project, and the decision to be made to continue the project. In most studies, the higher the amount invested in the study, the more likely the decision to continue on with the project, regardless of potential future gains. Using players from the first two rounds of the draft from 1980 to 1986, they found that teams were more likely to give higher draft picks more playing time, even after accounting for their statistical output on the court. The study found that draft pick order was a highly significant indicator of playing time, time until a trade, and length of career. Whether the relationship between draft pick order and the three variables is from teams giving their higher picks more time to show the abilities that the team thought the players had, or from top picks just being thought of as better because they were top picks, where a player was chosen in the draft was a significant factor through at least the first 5 years of a player's career. This makes it that much more important that those top picks be properly picked, since they historically will get more playing time and have longer careers than lower ranked players.

Taking what Staw and Hoang did and enhancing it, Camerer and Weber (1999) used an updated sample, the players taken in the first two rounds of the draft from 1986 through 1991, as long as the players had at least a two year career. This study went much more in depth, including additional factors not considered in the previous study, such as quality of a player's back up on his team, and rankings of a player before the draft. The rankings used by Camerer and Weber were pre-draft scouting reports, done by Don Leventhal, an NBA Analyst. These rankings differed from where a player was predicted to go in the draft. If a point guard may be very talented, but the top three teams all have All-Star caliber point guards, he will most likely go to the team picking fourth. In conclusion, Camerer and Weber found that there was an issue with sunk costs and overcommitting to high draft picks in the NBA, just as Staw and Hoang had said, but the strength of the relationship was only half the magnitude of Staw and Hoang's findings.

Flipping the script from sunk costs in draft picks, Groothuis, Hill, and Perri (2007) decided to find all of the superstars currently playing in the NBA, and see where those they were drafted. The authors define a superstar to be any player who had a player efficiency more than two standard deviations above the mean in each year from 1987 to 2004. This resulted in 12 to 22 players each year being considered “superstars,” at least by this measure. They also looked at the top five players each year in player efficiency. They found that there is no strong relationship between these players and draft pick, meaning that a superstar can come from anywhere in the draft, and often does. Beyond the #1 pick, which has produced many superstars, a lot of the top picks are “false positives”, players that do not develop into the type of player the teams that drafted them thought they would turn into.

This shows what has been obvious to everyone in the NBA: drafting is not an exact science. It would also do good to look at what coaches in the NBA value in their players. The All-Rookie NBA team consists of two teams of five players each, First Team and Second Team. Coaches vote for this award, which makes it unique among NBA awards, since most awards are voted on by the media. A coach cannot vote for his own player, so 29 coaches are available to vote for any given rookie. Coaches give two points for a player for a first team vote, and one point for a second team vote. This gives a possible range of 0 to 58 points for each player. When doing a regression using All-Rookie voting as the dependent variable, with points per game being the only explanatory variable, Berri, Brook, and Schmidt (2007) found that 74% of the variance in the voting can be explained. This shows that points per game is a huge factor for coaches in determining who will be All-Rookie. Even after adding in several additional counting stats, the coefficient of determination is only 76%, which means the additional variables do not have a big impact on the voting.

Considering the fact that scoring explains most of the variation in rookie of the year voting, and the voting is done only by the coaches, it appears that coaches are subject to overvaluing points scored, especially when points have never had an especially strong correlation with wins (Galletti, 2010). In recent years, there has been a heavier emphasis placed on scoring efficiency, and not just scoring. Organizations seem to be looking deeper into the data to figure out how to win games.

Since it has been established that draft position does have some bearing on a player's playing time and length of career, the next step is to look at how well college career statistics will predict where a player will be drafted. Coates and Oguntimein (2008) use points, rebounds, assists, steals, turnovers, blocks, personal fouls, field goal percentage, free throw percentage, and three formulas for productivity to predict draft position for all draftees between 1987 and 1989 who have at least one year experience in the NBA. The first productivity formula, used by the NBA as a comprehensive statistic for player productivity, is as follows:

(1) **((PTS+ REBS+ ASTS+ STLS+ BLKS)-(FGA-FGM) + (FTA-FTM) + (TO)) / (MP)**

Two additional formulas are used by Berri, Schmidt, and Brook (2006). The first does not account for assists, and the second does. They are as follows:

(2) **((PTS + REBS+ STLS+ (0.5) BLKS+ FGA- (0.5) FTA- TO- (0.5) PF)/ (MP)**

and

(3) **((PTS + REBS+ STLS+ (0.5) BLKS+ (0.5) ASTS - FGA- (0.5) FTA- TO- (0.5) PF)/ (MP)**

Coates and Oguntimein (2008) found that scoring, scoring efficiency (field goal percentage and free throw percentage), blocks, assists, and rebounds all have a positive correlation with a player’s draft stock, but only for schools from large conferences (ACC, Big 12, Big Ten, SEC, Pac 12, and Big East). For players from smaller conferences, scoring efficiency, free throw shooting, and blocks were the only factors that significantly improved their draft position. Scoring, rebounds, and assists did not have an effect on draft pick. This reveals that NBA teams don’t care about the raw totals of points, rebounds, and assists when it comes to players from small schools, since they are most likely playing against inferior competition. What they do look for is scoring efficiency, and the best way for a small school player to improve his draft stock is to improve his field goal percentage and free throw percentage.

Coates and Oguntimein performed a correlation analysis between college statistics and NBA statistics, using the same variables as before. Except for field goal percentage, every college statistic was significantly correlated with its corresponding NBA statistic. This proves that a relationship does exist between college performance and NBA performance, and therefore, we can try to quantify it to predict which players will be most successful. Breaking down the analysis, NBA points was used as the dependent variable in a regression analysis, with the same independent variables as in the previous analysis. There were four variables that were significant: college points per game, an interaction variable between college points per game and large conference/small conference, an interaction variable between college points per game and whether the player had played on a championship team, and draft pick.

When predicting minutes played, several variables were significant, including scoring and rebounding as the two most significant. But interestingly, college rebounds and NBA playing time were negatively correlated; meaning a higher rebounding average in college brought you less playing time in the NBA. There were additional interactions between college stats and NBA stats. For example, one additional rebound per game in college resulted in .25 more rebounds per game in the NBA. One additional steal per game in college led to .4 more steals as a pro. As for scoring, a 10% increase in scoring average in college led to the player having a 4% higher scoring average in the NBA. Perhaps the most telling relationship was between conference and length of NBA career: a player drafted from a large conference had, on average, a career that was four years longer than a player drafted from a small conference. This tells us that the school a player comes from is also an important factor.

Based on the aforementioned studies, it can be concluded that college statistics are indeed valuable and predictive when looking at NBA statistics, with points scored being the most significant variable in terms of draft position and All-Rookie voting. It can also be concluded that these statistics are not the only important information, or else choosing the most successful players would be much easier. With this limitation in mind, this paper will still attempt to quantify the likelihood of a college player having a successful NBA career.

Chapter 3: Background

History of the Draft

The first year of the draft was 1947, when the NBA was still the Basketball Association of America (BAA). The order of the draft was the inverse of the standings from the previous year. So the team with the worst record got the first pick, the team with the second worst record received the second pick, and so forth. This was in place until 1966, when the teams with the worst record in each conference would have a coin flip to see who received the #1 pick. The winner received 1st pick, the loser received 2nd, and the remaining teams went in the inverse of the standings, as it had been (NBA.com: Evolution of the Draft and Lottery, n.d.).

This stayed common practice until 1985, when the first lottery system was introduced. The purpose of the lottery was to discourage tanking, or the practice of intentionally losing games in order to garner a high draft pick. The Houston Rockets had received the #1 pick two years in a row, and owners were frustrated at the Rockets being guaranteed the #1 pick by having the worst record in the NBA. Under this new arrangement, all of the non-playoff teams would have an equal chance of landing the #1 pick. The first year, there were eight non-playoff teams, and all eight of their spots were selected through the lottery. Two years later, the NBA altered the draft rules so that only the first three picks of the draft would be chosen by the lottery, and the remaining teams would receive their draft picks by inverse order of the standings (NBA.com: Evolution of the Draft and Lottery, n.d.).

There have been small changes to the lottery system since its inception. In 1989, the NBA draft went down to only two rounds. In the past, teams were able to keep selecting players in later rounds until they ran out of prospects. The longest drafts were in 1960 and 1968, each of which lasted 21 rounds. In 1974, the NBA set the limit at 10 rounds. Eleven years later, in 1985, the length was shortened to seven rounds. Finally, in 1989, in agreement with the National Basketball Players Association (NBPA), the NBA set the draft to have two rounds. This shortened the duration of the draft, and limited the number of players who would be drafted. Any undrafted player would have the opportunity to try out with whichever team he wanted. The following year, in 1990, a weighted system was introduced, which gave the team with the worst record a much higher chance to win the lottery, and the non-playoff team with the best record a very small chance to win the lottery. This still gave the worst teams the best chance for the #1 pick, but gave all lottery teams a chance to win. As the number of teams in the NBA expanded, the number of teams included in the lottery increased. In 2004, the number of lottery teams was increased to 14, where it still stands today (NBA.com: Evolution of the Draft and Lottery, n.d.).

Draft Eligibility

There are certain requirements for players to be eligible for the draft. The rule created for NBA draft eligibility in the 1960s was the a player needed to complete 4 years of college before being eligible for the draft. Spencer Haywood, a star basketball player from Detroit, Michigan, attended Trinidad State Junior College in Colorado for one year, and subsequently transferred to the University of Detroit for an additional year. After a stellar performance at Detroit, where he average 32.1 points and 21.5 rebounds per game, Haywood decided to turn pro. But due to the NBA’s requirement of 4 years of college, Haywood was not eligible.

Therefore, he joined the Denver Rockets, who played in the American Basketball Association (ABA), a competing professional basketball league (NBA.com: Spencer Haywood Bio, n.d.).

Haywood continued his excellent play in the ABA. He averaged 30 points per game and 19.5 rebounds per game on his way to earning Rookie of the Year and MVP honors for the 1969-1970 season. Even though he was still not eligible for the NBA, Haywood joined the Seattle SuperSonics for the 1970-1971 season. The NBA once again said that he was not eligible to play, and in response, the owner of the SuperSonics filed an anti-trust lawsuit against the NBA, which turned into Haywood v. National Basketball Association. This case went all the way up to the U.S. Supreme Court, and in a 7-2 ruling, the Court stated that Haywood would be allowed to play. This launched a new policy by the NBA, stating that if a player could prove a financial hardship, they could be given special consideration to be eligible for the draft, even if they did not meet the criteria for eligibility (NBA.com: Spencer Haywood Bio, n.d.).

As far as clearing the way for high school players to join the NBA, this decision did not open the floodgates. In 1974 and 1975, a total of three players went straight from high school to professional basketball, one to the ABA and two to the NBA. After these players, it took 20 years for another high school player to be drafted. In 1995, Kevin Garnett was drafted fifth overall by the Minnesota Timberwolves. Garnett would go on to have one of the most successful careers ever, winning an NBA championship, All-Star and regular season MVP Awards, and garnering 15 All-Star game selections. The following year, Kobe Bryant was selected 13th overall by the Los Angeles Lakers. Bryant sits at third on the all-time scoring list, with five NBA championships and two NBA Finals MVP trophies on his resume. These success stories led to an influx of high school players being drafted, reaching a high point in the 2004 draft, where eight of the 29 first round picks were players out of high school (A look at High School Players…, n.d.). This led to a new Collective Bargaining Agreement (CBA) being negotiated between the players' union and the owners in 2005. The owners wanted to have a minimum age of 20 for draft eligibility, and the players did not want any limit. The final agreement was a minimum age of 19 years, which was only agreed upon after salary cap changes in the players' favor were put into place (NBA.com–CBA Principal Deal Points, n.d.).

The new CBA, negotiated in 2011, did not change the minimum age of eligibility for the draft, but it did call for the creation of a committee to discuss future changes. Thus, the current draft eligibility rules are as follows: First, all players must be at least 19 years of age during the calendar year of the draft. Second, if the player is from the United States, he must be at least 1 year removed from graduating high school. Most players choose to go to college and play basketball in the National Collegiate Athletics Association (NCAA), which is the governing body of all college sports. Others choose to play overseas, or join the Developmental League (D-League) of the NBA to prepare for life in the NBA (Table of Contents, n.d.).

The fact that for right now, and in the foreseeable future, high school players will not be eligible to be drafted makes this analysis even more important. With the overwhelming majority of players coming from the NCAA, being able to predict success that others might not be able to see is a distinct advantage.

PER and Win Shares

PER, or Player Efficiency Rating, is a statistic developed by John Hollinger, and it encompasses nearly every aspect of a player's production. The pertinent information about the rating is as follows: an unadjusted PER is calculated using a large number of variables, including points, rebounds, assists, field goals, free throws, turnovers, and three pointers, as well as team and league statistics. This variable is then adjusted for the team's pace, and normalized to have an average of 15 (Calculating PER, n.d.).

PER has a rough scale which demonstrates how a player has produced in a given year:

A Year for the Ages: 35.0
Runaway MVP Candidate: 30.0
Strong MVP Candidate: 27.5
Weak MVP Candidate: 25.0
Bona fide All-Star: 22.5
Borderline All-Star: 20.0
Solid 2nd Option: 18.0
3rd Banana: 16.0
Pretty good player: 15.0
In the rotation: 13.0
Scrounging for minutes: 11.0
Definitely renting: 9.0
The Next Stop: D-League: 5.0 (Player efficiency rating, 2014)

Since PER can be calculated for a single season or for multiple seasons, the scale provides a good guide as to just how productive a player has been over his career. As mentioned earlier, Michael Jordan is widely considered the greatest to ever play the game, and PER supports that claim. Jordan currently has the highest career PER at 27.91. LeBron James, the number one pick in the 2003 draft and four time MVP, is second, with a career PER of 27.76. There are only 62 players in the history of the NBA who have a career PER of 20.00 or higher.

Calculated by “Calculating Win Shares” (2014), win shares (WS) represents how much a player contributed to his team’s wins. Win shares is calculated as one game equaling roughly one win share, so a team that won 50 games in a season would have approximately 50 win shares in that season. The calculations do not come out exact, but the average absolute error for all teams since the 1962-63 season is 2.74 wins, with a root mean squared error of 3.41, which is relatively small compared to the 82 games. For example, the total win shares of all players on the Oklahoma Thunder in the 2013-2014 season is 59.4. The team actually won 59 games, so the absolute error for that team is 0.4. Therefore, win shares is a decent approximation of how much a player contributes to his team’s wins.

Sean Smith, a leader in baseball’s sabermetric movement, created a formula for win shares that was quite simple, using only points, missed field goals, missed free throws, rebounds, assists, blocks, steals, and turnovers. There are additional calculations to create marginal points and team points per win, but the formula is essentially very straightforward. This paper is using the Basketball Reference formula because it is more sophisticated, and all of the data has come from the site.

Win shares is divided into two categories: offensive win shares, and defensive win shares. Offensive win shares are calculated using points produced and offensive possessions. Points produced is a metric that captures how a player contributes to scoring, primarily through shooting, assists, and offensive rebounding (Points Produced Definition…, n.d.). Offensive possessions are predicted for each player. An offensive possession will end when (a) the team scores, (b) the team misses and the opponent gets the rebound, (c) the team turns the ball over, or (d) shooting free throws and either making the last shot or not securing the offensive rebound. Using these numbers from a game, the total number of possessions can be estimated for that game. This is then divided by two, and each team's estimated possessions are equal to the result. The ratio of a player's minutes to the total minutes for the team in a given season is multiplied by the team's estimated possessions to get a player's estimated possessions (Oliver, 2011, pp. 343-349).

These are used to calculate something called marginal offense. The formula is:

(4) *Marginal offense = (points produced) – 0.92*(league points per possession)*(offensive possessions)*

Another statistic, marginal points per win, is calculated using team and league pace. Pace is an estimate of how many possessions a team will have within 48 minutes, or the standard length of an NBA game. The league pace is an average of all teams. The formula is as follows:

(5) *Marginal points per win = 0.32*(league points per game)*((team pace)/(league pace))*

Finally, offensive win shares is the ratio of the two previous statistics:

(6) *Offensive win shares = (marginal offense)/(marginal points per win)*

Defensive win shares are calculated using Defensive Rating. Dr. Dean Oliver, an author and statistician, devised formulas for both points produced and defensive rating. The formula for defensive rating is:

(7) *Defensive rating = (Opponents points allowed/Opponent's possessions)*100*

The marginal defense and marginal points per win are calculated, similar to the offensive win shares. The final calculation for defensive win shares is:

(8) *Defensive win shares = (marginal defense)/(marginal points per win)*

To calculate win shares, offensive and defensive win shares are added together. A more detailed description of the formulas for win shares and PER can be found in Appendix A. For example, in 2013, Kevin Durant, the MVP of the league, accumulated 14.8 offensive win shares, and 4.4 defensive win shares, giving him 19.2 win shares for that season (Calculating Win Shares, 2014).

In addition to win shares, which is an accumulation statistic, there is win shares per 48 minutes (WSP48), which is how many wins a player contributes over the length of an NBA game. This is a rate statistic, and gives a better idea of how effective a player is on a per game basis, instead of how long a player's career was.

For example, Michael Jordan has the highest career win shares per 48 minutes, with .2505, but is fourth all time with 214.02 win shares. This shows that Jordan was very effective in his playing time, but did not have as long of a career as other superstars. Conversely, Kareem Abdul-Jabbar is the all-time leader in win shares with 273.41, but is seventh in win shares per 48 minutes at .2284. Since Abdul-Jabbar played 1,560 career games, and Jordan played 1,072, Abdul-Jabbar would have more opportunities to accumulate win shares, but his later years of his career brought down his win shares per 48 minutes. Over his final three seasons, Abdul-Jabbar averaged win shares per 48 minutes of .148, .111, and .082.

Chapter 4: Methodology

This paper will look at first round draft picks between 1985 and 2005. This time frame was chosen because the beginning of the current draft lottery began in 1985, and because most of the counting stats we use today (points, rebounds, assists, steals, blocks) were collected during college careers at this time. The endpoint of 2005 was chosen to allow players chosen in 2005 to have a long enough career (9 years) that we will have a good idea about how successful they have been in the NBA. College career statistics are used, including points per minute (PPM), rebounds per minute (RPM), assists per minute (APM), steals per minute (SPM), and blocks per minute (BPM). Additional information, such as All-American honors and player of the year awards will be included in the analysis as dummy variables to measure their significance.

Using PER as the primary dependent variable, a regression analysis of career PER on college statistics was performed. Regression on career PER using the rookie year statistics for each player was also conducted.

The big problem here is that PER does not necessarily translate into wins. According to “FAQ” (2014), the coefficient of determination between PER and team wins is only 33%, leaving two thirds of the variation in wins accounted for by different measurements. But one new relationship we can look at is between PER and win shares. PER correlates nicely with win shares, with a correlation coefficient of .699 for players in our sample. But the great relationship is between PER and win shares per 48 minutes, which calculates how many win shares a player contributes per 48 minutes of playing time. The correlation coefficient between these variables is .877, which is extremely useful. Therefore, this paper will also perform regression analyses with win shares as the dependent variable and with win shares per 48 minutes as our dependent variable. This will give a more valuable look, in terms of wins, at which players to draft.

Initially, this paper will discuss the relevance of college statistics to predicting NBA career success by doing a multiple linear regression on NBA career PER, WS, and WSP48 using college statistics. To validate the findings, PER, win shares, and win shares per 48 minutes of all of the players in our sample will be recorded and ranked to see where they should have been drafted, based on those values. Additionally, the model will be used to predict PER, win shares, and win shares per 48 minutes, and rank the players using those values. Finally, the analysis will compare the actual order of the draft, the ideal order of the draft, and our model's order of the draft, and see if the models achieved a better measure of future success.

This model can be tested using players drafted between 2006 and 2010, since they are outside of the range of the sample. This will be the validation sample. Testing these years in the validation sample will show how effectively or ineffectively teams are drafting in the NBA draft. Ideally, NBA teams would be able to use the results found here to better plan and prepare for upcoming drafts, knowing which variables are important and which are not.

The procedure was repeated, using rookie year statistics. After looking at the draft from 1985-2005, the players from 2006 draft and beyond were the validation sample, using the model to predict the PER, win shares, and win shares per 48 minutes of the top prospects, and ranking them accordingly. These rankings can be compared to the "ideal" rankings, where the first player chosen has the highest PER, second player chosen has the second highest PER, and so on. If the model does a better job predicting the success in the careers of these prospects, it can be conclude that there is a better way to prepare for the NBA draft.

Chapter 5: Data Description

The data collected for this study is very extensive. All first and second round draft picks between 1985 and 2005 were compiled, totaling 1152 observations. All data was collected from www.basektball-reference.com, an exhaustive database of basketball statistics. Each observation included 221 variables, including NBA total stats, per game stats, per 36 minute stats, per 100 possession stats, and advanced metrics. Those same stats were collected on the rookie year in the NBA, and the playoffs in the NBA. In addition, the NCAA total stats and per game stats are collected. To handle players on teams that did not make the NCAA tournament, a Tournament Seed of 17 was entered. This ensured that data was available for all players. There are also descriptive variables, such as flag variables for All-Rookie, All-NBA, and All-Defensive Teams, Rookie of the Year, Most Valuable Player, Defensive Player of the Year, Most Improved Player, Sixth Man of the Year, All-Star, All-Star MVP, Player of the Week, and Player of the Month. College descriptive variables include Made NCAA Tournament, Made Final Four, National Championship, AP All American Team, USBWA Player of the Year, AP Player of the Year, Helms Player of the Year, Naismith Player of the Year, Sporting News Player of the Year, UPI Player of the Year, NABC Defensive Player of the Year, NABC Player of the Year, Wooden Award Winner, USBWA Freshman of the Year, NCAA Tournament Most Outstanding Player, Rupp Player of the Year, and NIT Most Valuable Player. There is also a flag for McDonald's All American, a high school award.

Out of the 1152 total players, 163 did not play NCAA basketball. Of the 163, 39 came to the NBA directly out of high school, 115 were players playing in foreign countries, and 9 came from junior college or community college. Also out of the 1152 total players, 162 have never played in the NBA. Since our sample requires NBA players with NCAA experience, players who played internationally, went straight from high school to the NBA, or have never played in the NBA were eliminated. Considering some players would have neither NCAA experience nor NBA experience, there is a bit of overlap in these groups. Finally, players who played at the Division II or Division III level were removed. This resulted in 841 players who had both an NCAA Division I and an NBA career.

There are some obvious differences between the two rounds right away. Before the filtering, there were 575 first round picks and 577 second round picks. After stripping away those without NBA or NCAA experience, the sample consists of 487 first round picks and 354 second round picks. As has been proven by other studies, a first round pick has a much higher chance of having an NBA career than a second round pick.

Getting into the numbers, the average PER for all players in our sample is 12.29 with a standard deviation of 5.16. Dividing into the two rounds, the first round has an average PER of 13.43 with a standard deviation of 3.67, while the second round has an average PER of 10.71 with a standard deviation of 6.31. This shows that the first round has a significantly higher PER, with much less variability than the second round. Using the medians tells a similar story, with the first round picks having a median PER of 13.6, and the second round having a median PER of 10.85.

Table 1

Summary Statistics for PER, Win Shares, and Win Shares per 48 Minutes

	N	Mean	Standard Deviation	Median	Min	Max
WHOLE SAMPLE						
PER	841	12.29	5.16	12.40	-22.1	76.1
Win Shares	841	21.97	30.63	9.4	-1.6	234.6
Win Shares per 48 minutes	841	.062	.098	.069	-.638	1.442
FIRST ROUND						
	N	Mean	Standard Deviation	Median	Min	Max
PER	487	13.43	3.67	13.6	-4.5	26.4
Win Shares	487	31.39	34.91	21.3	-1.6	234.6
Win Shares per 48 minutes	487	.075	.056	.081	-.326	.250
SECOND ROUND						
	N	Mean	Standard Deviation	Median	Min	Max
PER	354	10.71	6.38	10.85	-22.1	76.1
Win Shares	354	9.01	16.23	1.00	-1.3	108.9
Win Shares per 48 minutes	354	.043	.134	.047	-.638	1.442

Whereas PER follows a relatively normal distribution, WS has a right skewed distribution. This makes sense when the process of how each measurement is calculated is considered. PER is a normalized variable with 15 being the league average by design. WS is an accumulated variable, and while players can have negative WS, most teams would not hold onto a player long enough for them to collect a large amount of negative win shares. Therefore, WS is practically left censored at 0, and is not limited on the right (see Appendix B).

Once again, it is shown that the first round players out-produce second round players. This time, the first round has a higher standard deviation in WS, which means it is much more volatile than the second round. This is due to the distribution being right skewed, and the fact that no second round pick in our sample has had more than 108.9 WS. So in this case, the variability is a positive aspect. It is also interesting to note that the median WS for a second round player who played in the NBA is 1.00, which is an absolutely miniscule amount.

As for WSP48, we can see that it follows a relatively normal distribution; similar to PER (See Appendix B). The variability for WSP48 is also much higher for second round picks than first round picks. This whole image shows that first round picks are more likely to have higher PER's, WS, and WSP48 than second round picks, and with less variability for PER and WSP48.

Chapter 6: Analysis

College Statistics

The first analysis completed for this study was looking at college statistics regressed against NBA career PER. Certain variables were not available for a large amount of players, and have been subsequently eliminated from the analysis. Those include three pointers (made, attempted, percentage), minutes (total and per game), and turnovers and personal fouls (total and per game). Table 2 shows the results from that analysis.

These results were compiled using a forward selection method, whereby each variable is added to the formula, and if, at that time, the variable's p-value is not less than 0.15, it will not be brought into the equation. For PER, the formula ended up with 10 significant variables, and WS and WSP48 had 14 and 6 significant variables, respectively.

The analysis shows that there are several significant variables in predicting success at the NBA level. On all three regressions (PER, WS, and WSP48), field goal percentage was significant and the coefficient was positive. This shows that an increase in college field goal percentage leads to a higher PER, WS, and WSP48 in the NBA. It is also interesting that a player's team's tournament seed was significant in each of the regressions, supporting the idea that great players will make their teams great. It also makes sense that the best teams recruit the best players, and will therefore have the higher seeds in the tournament.

Table 2

Regression of College Variables against NBA PER, Win Shares, and Win Shares per 48 Minutes, Full Sample (N = 813)

Variable	**PER**	**Win Shares**	**Win Shares per 48 Minutes**
Intercept	3.126(2.016)	-51.81(11.054)	-0.1096(0.0511)
Field Goals Made	0(0.006)		
Field Goals Attempted		0.021***(0.007)	
Field Goal Percentage	6.871*(3.715)	67.229***(19.93)	0.1625**(0.0689)
Free Throw Percentage			0.064(0.0417)
Rebounds			0.0001**(0.0001)
Assists	0.013***(0.004)		0.0003***(0.0001)
Steals		-0.065(0.183)	
Blocks		0.142***(0.04)	0.0003*(0.0001)
Points Per Game	0.149*(0.081)		
Rebounds Per Game	0.14(0.097)	1.154**(0.53)	
Assists Per Game		2.969***(0.782)	
Steals Per Game	0.767**(0.347)	11.641*(5.934)	
Blocks Per Game	0.627***(0.225)		
Tournament Seed	-0.102***(0.033)	-0.467**(0.182)	-0.0011*(0.0006)
1st Team All-American	1.113**(0.5)	13.079***(3.082)	
AP Player of the Year		34.912***(9.054)	
NABC Defensive Player of the Year		10.723**(4.865)	
NABC Player of the Year		-20.179**(8.373)	
USBWA Freshman of the Year	2.114(1.374)	15.07*(7.678)	
NIT Most Valuable Player		17.074*(9.569)	
Summary Statistics			
R^2	.117	.257	.050

Standard errors are given in parentheses next to the coefficients. The coefficients are significant at the *10%, **5%, and ***1% levels.

One of the biggest trends was that the per-game results were very significant for the PER regression, whereas the counting totals were significant for the WSP48. It could have been guessed that the per-game stats would have been significant for the WSP48 as well, since it not a cumulative statistic, but a rate statistic.

Since there has been such a marked difference between the first and second round picks, especially in the standard deviation for PER, WS, and WSP48, it would be beneficial to split the dataset into the first and second rounds and repeat the same regression analysis. The tables are located in Appendix C.

When broken down by round, it is evident that the first round is much easier to predict. Once again, field goal percentage is significant on all three analyses, and has a positive coefficient. Also, blocks and 1st team All-American are significant across all three. This matches with the findings of Coates and Oguntimein (2008), who found scoring efficiency and blocks to have a positive correlation on predicting draft position. There are several players drafted very highly that most likely contribute to these positive correlations, such as David Robinson (1987 draft, #1 pick), Shaquille O'Neal (1992, #1), and Tim Duncan (1997, #1). These were all 7 footers who shot very high percentages and had high block totals in college, also all making 1st team All-American at least once. All of these players accumulated a PER of at least 24, total win shares of 175 or more, and win shares per 48 of at least .200, all of which are extremely high. These players help skew the significance of blocks, field goal percentage, and 1st team All-American.

The first analysis with PER as the dependent variable shows that steals, blocks, rebounds per game, assists per game, 1st team All-American, AP Player of the Year, and USBWA Freshman of the Year are significant, and all have positive coefficients. Coates and Oguntimein (2008) also found blocks, assists, and rebounds per game positive and significant in their analysis for predicting draft position. Since their sample was contained inside of the one used in this paper, this is an encouraging sign. No team statistics were significant at any level, implying that a player's success can be predicted solely with their numbers in college.

As for the win shares analysis, 1st team All-American, AP Player of the Year, NABC Defensive Player of the Year, and NABC Player of the Year are significant, and all but the NABC Player of the Year are positive. This could be attributed to Shawn Respert and Jay Williams, winners of the award in 1995 and 2002, respectively. These players won the NABC POY award, but not many other prestigious awards, such as AP POY. Both players had very disappointing careers, with neither accumulating a PER above 13, more than 2.5 win shares, or win shares per 48 minutes above 0.050.

In the third regression, the only accolade that is significant is 1st team All-American. The remaining variables are field goals made, field goal percentage, points, rebounds, assists, blocks, and steals per game, all of which have positive coefficients with the exception of field goals made. This equation actually had a significantly lower R^2 than the other two, with only 11.49% of the variance in NBA WSP48 being explained by college variables.

With such diversity in the statistics between all of the different positions in basketball, it seemed beneficial to separate the 1st round by position, and create a regression analysis for each data set. The results from this are located in Appendix D, Tables C through G. Here, a similar pattern in terms of coefficients of determination can be seen, as they range from 0.139 to 0.222 for point guards, shooting guards, and small forwards. But for power forwards, the equation for WS has an R^2 of 0.385, significantly higher than the other positions. Centers, however, show the greatest improvement. The WS equation for centers has an R^2 of 0.645, or nearly two-thirds of the variance. This is a drastic improvement, and shows the centers and power forwards are considerably easier to predict than point guards, shooting guards, or small forwards.

Point guards had very few variables that were actually significant. In the WS equation, for example, the only two variables are rebounds (0.207 parameter estimate), and assists per game (7.113). WSP48 isn't much better, with only 2nd team All-American (-0.048) and 3rd team All-American (-0.0416) being significant. No individual statistics are relevant to predicting NBA success. This leads to the conclusion that point guards are very difficult to project, and there aren't many differentiating college factors between good NBA point guards and bad NBA point guards.

As for shooting guards, the only variable to be significant across all three equations was assists per game. All coefficients were positive, which means that a shooting guard who is adept at passing the ball has a better chance of becoming a productive player in the NBA. Assists are also relevant for small forwards, as that is the only variable to be significant across all three equations at that position. Once again, all coefficients are positive. The general takeaway is that when dealing with either a shooting guard or a small forward, assists is the most indicative factor of future success.

The one variable significant for all three equations for power forward is AP Player of the Year, which is positive for all three. NABC Defensive Player of the Year is significant for WS and WSP48, which shows that, historically, the best power forwards in the country in the NCAA will go on to have success in the NBA. Another interesting item to note is that this was the first position where team record became significant. Team Losses and Conference

Winning Percentage are significant for PER and WS. No other position had been influenced by team success. This could mean that good power forwards make their teams better, or just that good power forwards have come from historically successful teams.

Weight is a very significant variable for centers, as all three equations have negative coefficients that are significant at the 0.05 level. Historically, heavier centers have performed worse than lighter centers. There are obvious exceptions to this (Shaquille O'Neal at 325 pounds is an obvious example), but the overall trend has been towards lighter centers performing better. This could be due to a stamina issue, as the heavier a player is, the more likely he is to wear down faster.

For the predicting of draft order, both equations (1st round as a whole, and 1st round broken out by position) will be used to see which one does the better job.

The analysis for second round shows why most teams consider second round picks to be like lottery tickets. There are only three significant variables among all three analyses, all in the WS equation. Additionally, none of the regression equations are significant, even at α=.10. In light of these facts, the remaining analysis will look at the first round picks only.

Rookie Year Statistics

The first round of regression analysis showed that there is a relationship between a player's college career and their professional career. But the amount of variance determined by the college statistics was not very high. In fact, other than the equations for power forwards and centers, the highest R^2 of any of the regression analyses was .222, or just over one fifth of the variance explained. So, some of what makes a successful NBA player can be determined, but can any information after the first year of NBA experience be gained?

Using the same methods as the regression with college variables, the formula used many available metrics from a player's rookie year, including games, minutes (total and per game), points, rebounds, assists, steals, blocks, turnovers, personal fouls (total, per game, and per 36 minutes), field goals, three pointers, free throws (made, attempted, percentage), PER, Win Shares (offensive, defensive, total), Win Shares per 48 minutes, Rookie of the Month, All-Rookie 1st team and All-Rookie 2nd team. The results are included in Appendix E, Tables H through M.

There has been a plethora of information gained after one year in the NBA. Starting with the PER regression, weight, draft position, games, free throw percentage, steals, personal fouls, personal fouls per game, steals per 36 minutes, PER, offensive win shares, and defensive win shares are all significant at some level, with draft position, games, steals, personal fouls, steals per 36 minutes, PER, and defensive win shares significant at the 1% level. This shows that a player who is a high draft pick, plays a large amount of games, and has a high PER in their rookie season will probably have a higher career PER. Figure 1 shows the relationship between rookie year PER and career PER. This may not be earth-shattering news, but it does give insight into what GM's should be looking at when evaluating potential trades. What does help as well is that the R^2 has increased to .634, compared to .207 using only college variables.

The vast outlier in Figure 1 is courtesy of Yinka Dare, a center drafted 14th in 1994. Dare played only one game in his rookie season, and was in that game for three minutes. He missed his only shot attempt, and committed one turnover and two fouls to go along with one rebound. This led to his rookie year PER to be an astronomically bad -33.9, which drastically skews the graph.

Similarly, an increase in R^2 for the WS regression is observed. For using college stats, $R^2 = .213$, and using rookie year stats, $R^2 = .487$. In this analysis, there was draft position, games, personal fouls, turnovers per game, defensive win shares, and win shares being significant at some level. Draft position, personal fouls, turnovers per game, and win shares are significant at the 1% level.

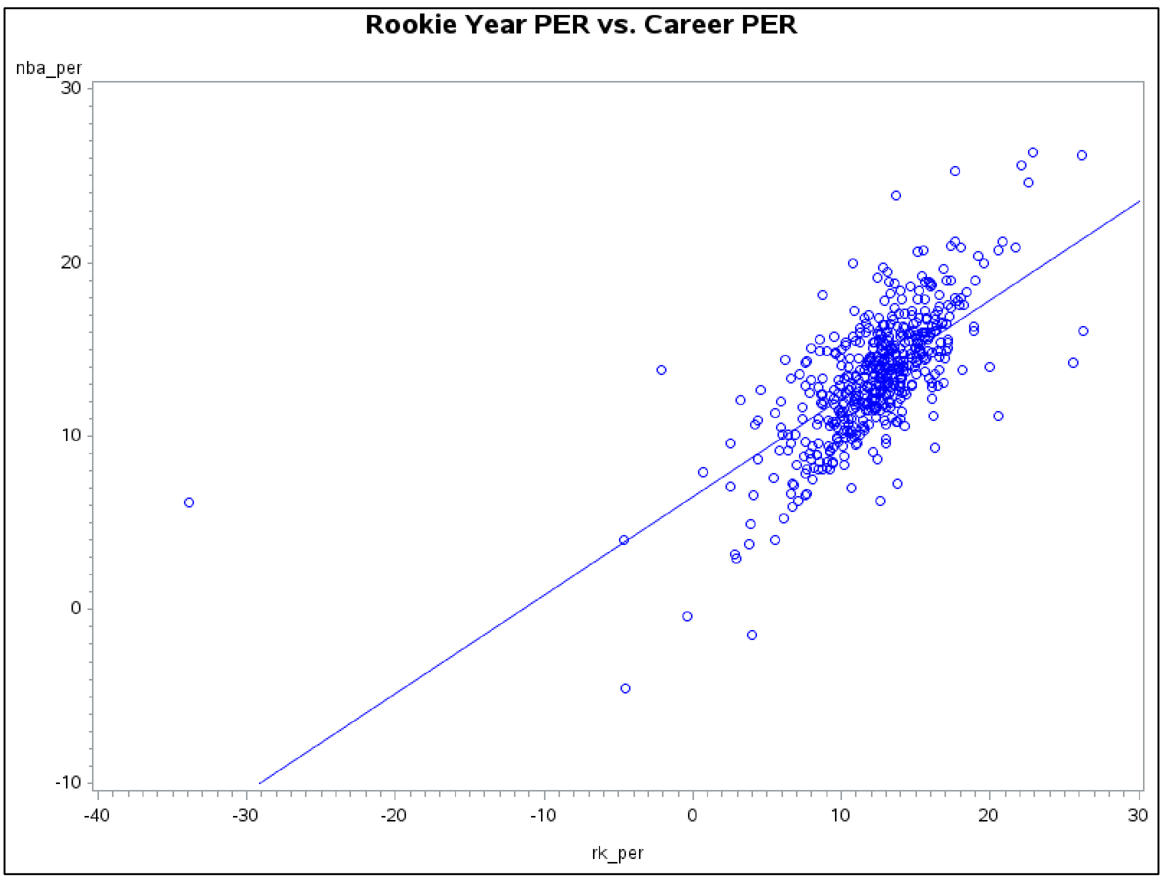


Figure 1

Scatterplot of Rookie Year PER against Career PER with Regression Line

In the WSP48 regression, a large amount of information is gained as well. The R^2 goes from .1149 with college stats to .559 with rookie year stats. Significant variables include draft position, games, free throw percentage, personal fouls, steals per game, personal fouls per game, steals per 36 minutes, win shares, and win shares per 48 minutes. With such a high R^2, these variables tell quite a bit about how efficient a player will be throughout their career. The relationship between rookie WSP48 and career WSP48 is shown in Figure 2.

Similar to Figure 1, there is an extreme outlier; or in fact two, in this case. The most extreme is once again attributed to Yinka Dare, courtesy of his 3-minute-long rookie season. His WSP48 that season was -0.736. The second player was Troy Bell, the 16th pick in 2003. Bell played in six games, totaling 34 minutes. His abysmal shooting was the source of the low WSP48, as he was 4-18 from the field (22.2%), and 0-4 on three point attempts. This, combined with relatively high rates of fouls and turnovers, led to a WSP48 of -0.326. Without these two individuals, the relationship looks much stronger.

The analysis by position is even better, as far as R^2 goes. For the center position equation for WS, the coefficient of determination is 0.824. This agrees with the college variables in that centers are the easiest to predict success. Conversely, the point guard WS regression equation has a coefficient of determination of only 0.405. Predicting success for point guards does not get much easier, even with a season of experience to analyze.

Due to the fact that PER was not necessarily correlated strongly with wins, and since wins have a higher correlation to WSP48, this paper will be using WSP48 as the primary variable in the following section of analysis.

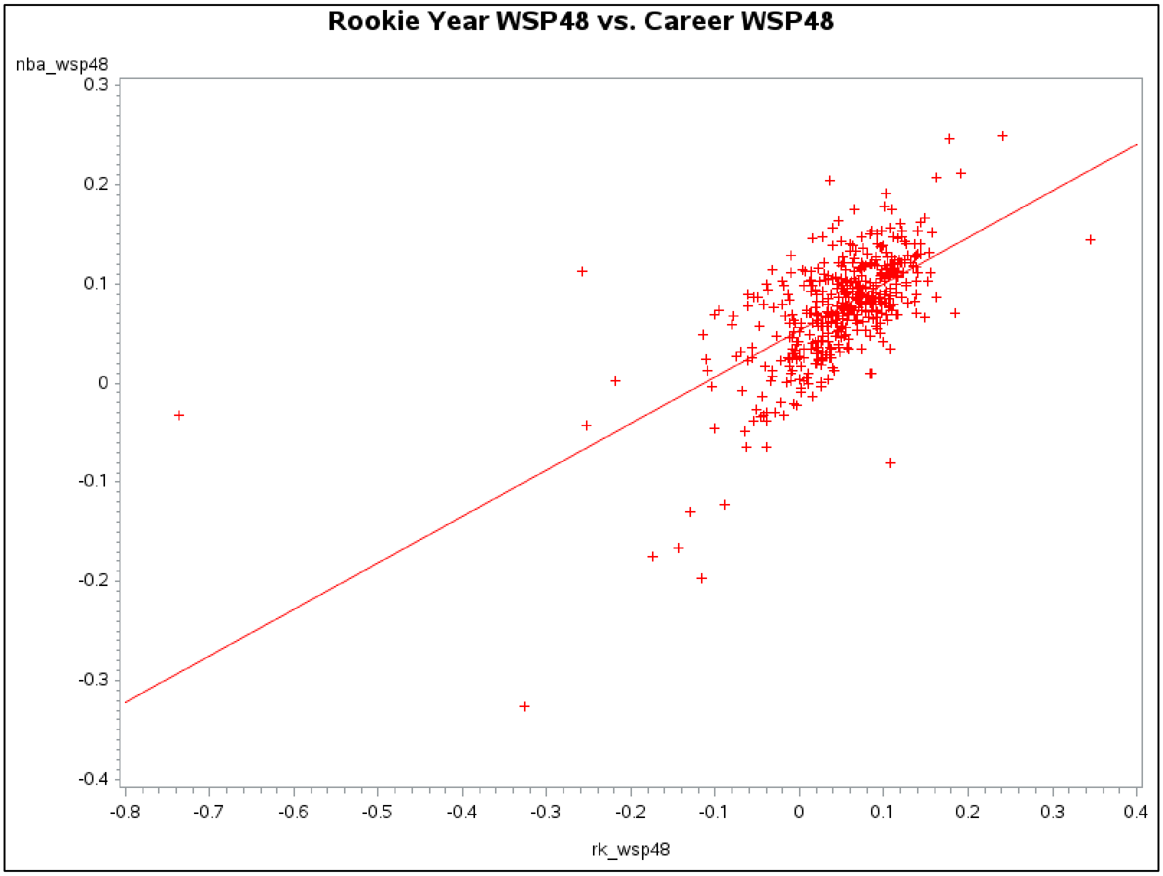


Figure 2

Scatterplot of Rookie Year WSP48 against Career WSP48 with Regression Line

Predicting Draft Order

This is the most interesting part of the whole process. Using all variables that were significant at the 10% level, the paper predicted a player's WSP48 with both college variables and rookie year variables. From the perspective of a GM, the ability to predict a player's success before he is drafted is extremely important, but the ability to re-evaluate all players after one year in the NBA to determine which players are most likely to be successful is just as crucial. This will determine what players they pursue in free agency and through trades.

The first analysis will look at re-ranking each draft based on the predicted value of WSP48 using college variables. The model with all 1st round players will be referred to as Model 1, or M1, and the model split apart by position will be referred to as Model 2, or M2. Since exact predictions are not possible, this paper focused on ranking the players in relation to one another, comparing the predicted order of the draft with the ideal order. The ideal order is the ranking from largest to smallest of career WSP48. The 1985 draft is shown in Table 3 to illustrate this point.

Another issue was dealing with foreign and high school players. The decision was made to not use a player's actual draft pick, but where they were drafted out of those who attended college. This made it easy to compare all three rankings.

Some pretty interesting things are observed from this table. The biggest question was how well the formula would predict Karl Malone, who was chosen 13th overall, but ended up with one of the best careers of all time. Unfortunately, the M1 ranking had him 11th out of players attending college, which is only one spot above where he was chosen in the actual draft, and M2 had him ranked 8th. But the equations did predict correctly that A.C. Green should have been chosen higher than the last pick of the draft. His actual WSP48 rank is 5th, and his M1 and M2 predicted ranks were 7th and 6th, respectively, much better than the 22nd where he was actually selected.

Table 3

1985 NBA Draft with Ideal Ranking and Predicted Ranking

Name	Draft Pick	NBA WSP48	Ideal Rank	Pred. WSP48 (M1)	Rank (M1)	Pred. WSP48 (M2)	Rank (M2)
Ewing, Patrick	1	**.150**	3	.163	1	.177	1
Tisdale, Wayman	2	**.092**	8	.115	2	.088	9
Benjamin, Benoit	3	**.072**	14	.112	3	.108	4
McDaniel, Xavier	4	**.091**	9	.089	6	.129	2
Koncak, Jon	5	**.085**	11	.085	8	.086	10
Kleine, Joe	6	**.062**	17	.067	14	.061	20
Mullin, Chris	7	**.139**	4	.091	5	.095	7
Schrempf, Detlef	8	**.156**	2	.070	10	.125	3
Pinckney, Ed	9	**.130**	6	.112	4	.104	5
Lee, Keith	10	**.062**	16	.069	13	.064	19
Green, Kenny	11	**-.022**	21	.047	20	.065	18
Malone, Karl	12	**.205**	1	.069	11	.088	8
Hughes, Alfredrick	13	**-.029**	22	.023	22	.034	22
Rasmussen, Blair	14	**.080**	12	.046	21	.069	16
Wennington, Bill	15	**.087**	10	.079	9	.076	14
Blab, Uwe	16	**.009**	20	.062	16	.079	12
Dumars, Joe	17	**.118**	7	.057	17	.073	15
Harris, Steve	18	**.053**	19	.067	15	.069	17
Vincent, Sam	19	**.069**	15	.055	18	.046	21
Catledge, Terry	20	**.074**	13	.054	19	.080	11
Reynolds, Jerry	21	**.060**	18	.069	12	.077	13
Green, A.C.	22	**.131**	5	.086	7	.097	6

Naturally, there must be a way to determine if the predicted draft order is better than the actual draft order. The initial attempt was to look at the absolute value of the difference between the predicted and ideal pick, and compare that to the absolute value of the difference between the actual and ideal pick. These would then be totaled for each draft, and the smallest total would be the most effective order. In addition, the differences were squared, and added up as well. This gives two good measurements of the effectiveness of the predicting. The first measurement does not punish for a large miss, but the second measurement does. For example, if a player's ideal pick was 10, they were actually picked 5th, but the predicted pick was 20th, the absolute value of the differences would be 5 and 10, respectively. This would not be a large difference. But the squared differences would be 25 and 100, resulting in a much larger disparity. So the second measurement will be looked at to see how many "large" misses occurred.

This also allows the calculation of Spearman's rank correlation coefficient, which is a measurement of dependence between two variables. It is a non-parametric measurement, which means it does not require any assumptions about our data. The resulting value will be between -1 and 1, which will tell us the magnitude and the direction of the relationship between the variables. Using our sample size per each year, n, and the actual and predicted values, the Spearman coefficient can be calculated with the following formula:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}. \qquad (9)$$

d_i^2 is the difference between the two observations. In this case, there were three sets of data to calculate the Spearman coefficient: the actual vs. the ideal, the M1 predicted vs. the ideal, and the M2 predicted vs. the ideal.

Table 4 shows the differences and squared differences between the actual, ideal, and predicted ranks, as well as the Spearman coefficient for each ranking.

The results of the first year in the study are very promising, with the M1 predicted rank having a sum of error of 98 and the M2 predicted rank totaling 82, while the actual draft had a sum of error of 142. The same held true for the squared differences. This is illustrated by the six picks that were at least 10 spots away from where the ideal ranking was for that player, compared to only three such picks for the M1 predicted ranks, and one for M2. There was also one predicted rank that matched the ideal ranks exactly for each model. Table 5 shows the summary of the rest of the years in our sample. Overall, the model rankings did a better job predicting, in winning more by years, by total differences, and with higher Spearman coefficient's in more years, even though the model's R^2 was only .1149. M1 had a higher Spearman coefficient in 14 out of the 21 years in the sample, whereas M2 had a higher coefficient in only 12 of 21 years. The sum of the years for M1 was the lowest out of all three, meaning that the equation from all 1st round players was more effective at predicting WSP48 than Model 2.

Table 4

1985 Draft–Absolute Value of Differences and Differences Squared of Ideal Rank and Model Ranks

Name	Draft Pick (d)	Ideal Rank (i)	ABS of (d-i)	$(d-i)^2$	**M1**	ABS of (m1-i)	$(m-i)^2$	**M2**	ABS of (m2-i)	$(m2-i)^2$
Ewing, Patrick	1	**3**	2	4	**1**	2	4	**1**	2	4
Tisdale, Wayman	2	**8**	6	36	**2**	6	36	**9**	1	1
Benjamin, Benoit	3	**14**	11	121	**3**	11	121	**4**	10	100
McDaniel, Xavier	4	**9**	5	25	**6**	3	9	**2**	7	49
Koncak, Jon	5	**11**	6	36	**8**	3	9	**10**	1	1
Kleine, Joe	6	**17**	11	121	**14**	3	9	**20**	3	9
Mullin, Chris	7	**4**	3	9	**5**	1	1	**7**	3	9
Schrempf, Detlef	8	**2**	6	36	**10**	8	64	**3**	1	1
Pinckney, Ed	9	**6**	3	9	**4**	2	4	**5**	1	1
Lee, Keith	10	**16**	6	36	**13**	3	9	**19**	3	9
Green, Kenny	11	**21**	10	100	**20**	1	1	**18**	3	9
Malone, Karl	12	**1**	11	121	**11**	10	100	**8**	7	49
Hughes, Alfredrick	13	**22**	9	81	**22**	0	0	**22**	0	0
Rasmussen, Blair	14	**12**	2	4	**21**	9	81	**16**	4	16
Wennington, Bill	15	**10**	5	25	**9**	1	1	**14**	4	16
Blab, Uwe	16	**20**	4	16	**16**	4	16	**12**	8	64
Dumars, Joe	17	**7**	10	100	**17**	10	100	**15**	8	64
Harris, Steve	18	**19**	1	1	**15**	4	16	**17**	2	4
Vincent, Sam	19	**15**	4	16	**18**	3	9	**21**	6	36
Catledge, Terry	20	**13**	7	49	**19**	6	36	**11**	2	4
Reynolds, Jerry	21	**18**	3	9	**12**	6	36	**13**	5	25
Green, A.C.	22	**5**	17	289	**7**	2	4	**6**	1	1
Sum			142	1244		98	666		82	472
Spearman coefficient				0.298			0.624			0.733

Table 5

Summary of Differences and Squared Differences between Ideal, Actual, and Predicted Rankings of WSP48, by Year (College Statistics)

Year	Sum of $[ABS(d-i)]^2$ with Spearman	Sum of $[ABS(m1-i)]^2$ with Spearman	Sum of $[ABS(m2-i)]^2$ with Spearman
1985	1244 (0.298)	666 (0.624)	472 (0.733)
1986	1374 (0.108)	850 (0.448)	1166 (0.243)
1987	1016 (0.426)	1400 (0.209)	1330 (0.249)
1988	910 (0.650)	1840 (0.292)	1514 (0.418)
1989	1772 (0.318)	1841 (0.292)	1786 (0.313)
1990	2381 (0.273)	1690 (0.484)	1885 (0.425)
1991	1478 (0.432)	1482 (0.430)	1802 (0.307)
1992	1674 (0.489)	1552 (0.526)	1392 (0.575)
1993	2992 (0.087)	2322 (0.291)	2078 (0.366)
1994	2270 (0.307)	3090 (0.057)	2454 (0.251)
1995	3268 (0.106)	2518 (0.311)	2322 (0.365)
1996	1084 (0.388)	1156 (0.347)	1424 (0.196)
1997	1292 (0.558)	1014 (0.653)	2076 (0.290)
1998	1336 (0.419)	1120 (0.513)	902 (0.608)
1999	836 (0.528)	764 (0.569)	1494 (0.156)
2000	1760 (0.130)	1032 (0.490)	1192 (0.411)
2001	898 (0.212)	694 (0.391)	482 (0.577)
2002	964 (0.275)	846 (0.364)	536 (0.597)
2003	522 (0.360)	428 (0.475)	228 (0.721)
2004	488 (0.129)	574 (-0.025)	458 (0.182)
2005	1638 (0.191)	1214 (0.400)	1936 (0.043)
Total SS	**31197**	**28093**	**28929**

To validate these findings, the formula was applied to players from the first round of the draft between 2006 and 2010. Since most of these players are still in the NBA and thus, their careers are not over, this validation is not ideal, but it will still give an idea if the predictions are holding up over time. Results are located in Table 6. Similar results occur in our validation sample. All 5 years have a higher Spearman coefficient for the M1 than the actual ranking. M2, however, does a terrible job, and only has a higher Spearman coefficient in one of the five years. In three out of the five years, the Spearman coefficient is greater than 0.5 for M1, indicating that the predicted rank has a strong correlation with the ideal rank. It is amazing how well this model works in predicting a draft's ranking by WSP48, given its low R^2.

Given the fairly low coefficients of determination, it is surprising that the model does a superior job of predicting rankings of the draft. Since there has been this level of success, it can be expected that the rankings using the rookie year information will be much closer to the ideal rankings than the actual rankings.

Table 6

Summary of Differences and Squared Differences between Ideal, Actual, and Predicted Rankings of WSP48 in Validation Sample, by Year (College Statistics)

Year	Sum of $[ABS(d-i)]^2$ with Spearman	Sum of $[ABS(m1-i)]^2$ with Spearman	Sum of $[ABS(m2-i)]^2$ with Spearman
2006	1600 (0.304)	1040 (0.548)	1550 (0.326)
2007	1326 (0.490)	1083 (0.583)	1964 (0.245)
2008	2648 (0.095)	2134 (0.270)	3040 (-0.039)
2009	1808 (0.214)	978 (0.575)	2158 (0.062)
2010	2774 (0.317)	2054 (0.494)	3188 (0.215)
Total SS	**10156**	**7289**	**11900**

Again, the formulas will be predicting the rankings of the draft using WSP48 for the 1985 draft through the 2005 draft, and will be validating those results with the 2006 through 2010 drafts. The results using WSP48 are located in Table 7.

Here, the rookie year model does a significantly better job than how the actual draft picks went. M1 performed better in 20 of the 21 years, and M2 had a higher coefficient for every single year. On top of that, M2's coefficient was higher than M1 in 15 of the 21 years, implying that the model by position does a much better job of predicting success after a player's rookie year. The Spearman coefficients were also incredibly high for some years. In 1987, the coefficient for M2 was 0.932, nearly a perfect correlation. For the 2003 draft, the Spearman coefficient for M1 was .944. The biggest difference was for Dwayne Wade, who should have been the first college player taken off the board, but was predicted 5th by the model. With the exception of Wade and Mike Sweetney, who is 7th in WSP48 but predicted 4th by the model, all players were predicted within two spots of their ideal rank. This is absolutely crucial information for GM's to be able to determine which players they will want to acquire, and which players they will want to avoid. It will also help prevent a team from holding on to a player just because they invested a high draft pick into them. If there is evidence that a player will not be successful, a franchise needs to move on from them, regardless of what they initially thought they would be getting from that player.

Additionally, the M1 ranking is the best in all 5 years of the validation sample for WSP48 (Table 8). Each of the Spearman coefficients are greater than 0.5 for M1, with four of the coefficients north of 0.7. M2 does a good job as well, but does not have a higher coefficient than M1 in any of the years. However, M2 does have a higher coefficient than the actual draft order in every year.

Table 7

Summary of Differences and Squared Differences between Ideal, Actual, and Predicted Rankings of WSP48, by Year (Rookie Year Statistics)

Year	$[ABS(d-i)]^2$ with Spearman	$[ABS(m1-i)]^2$ with Spearman	$[ABS(m2-i)]^2$ with Spearman
1985	1244 (0.298)	1072 (0.395)	550 (0.689)
1986	1374 (0.108)	604 (0.608)	612 (0.602)
1987	1016 (0.426)	282 (0.841)	120 (0.932)
1988	910 (0.650)	902 (0.653)	810 (0.688)
1989	1772 (0.318)	1328 (0.489)	938 (0.639)
1990	2381 (0.273)	825 (0.748)	1000 (0.695)
1991	1478 (0.432)	674 (0.741)	618 (0.762)
1992	1674 (0.489)	1856 (0.433)	1452 (0.557)
1993	2992 (0.087)	1548 (0.527)	1210 (0.631)
1994	2270 (0.307)	1438 (0.561)	1242 (0.621)
1995	3268 (0.106)	860 (0.765)	888 (0.757)
1996	1084 (0.388)	884 (0.501)	562 (0.683)
1997	1292 (0.558)	1138 (0.611)	812 (0.722)
1998	1336 (0.419)	786 (0.658)	582 (0.747)
1999	836 (0.528)	242 (0.863)	310 (0.825)
2000	1760 (0.130)	934 (0.539)	1016 (0.498)
2001	898 (0.212)	332 (0.709)	154 (0.865)
2002	964 (0.275)	412 (0.690)	320 (0.759)
2003	522 (0.360)	46 (0.944)	108 (0.868)
2004	488 (0.129)	292 (0.479)	192 (0.657)
2005	1638 (0.191)	618 (0.695)	586 (0.710)
Total SS	**31197**	**17073**	**14082**

With the knowledge that the validation sample supports the formulas from our sample, it can be concluded with some certainty that these models will be successful in the future, even with the changing landscape of the NBA. Granted, this is using data after the draft to re-predict the draft, essentially. So the following section discusses the effect of predicting the draft ranking of the top prospects of the NBA draft, without knowing where they actually went.

Table 8

Summary of Differences and Squared Differences between Ideal, Actual, and Predicted Rankings of PER, WS, and WSP48, by year (Rookie Year Statistics)

Year	Sum of $[ABS(d-i)]^2$ with Spearman	Sum of $[ABS(m1-i)]^2$ with Spearman	Sum of $[ABS(m2-i)]^2$ with Spearman
2006	1600 (0.304)	462 (0.799)	594 (0.742)
2007	1326 (0.490)	576 (0.778)	836 (0.678)
2008	2648 (0.095)	1328 (0.546)	1954 (0.332)
2009	1808 (0.214)	682 (0.703)	726 (0.684)
2010	2774 (0.317)	966 (0.762)	1312 (0.677)
Total SS	**10156**	**4014**	**5422**

Predicting All Draftees

Since it is not known who will be a first round pick before the draft, there is no way of just taking the first round picks and arranging them based on the model. So the logical next step would be to include all of the top prospects heading into the draft, rank them with the models, and compare that to how they all fared.

The year 2009 was chosen because that was the first year that data is available for the top 100 prospects going into the draft. There is a “Big Board” of the top 100 prospects, both college and international, for the 2009 NBA draft (Smith, 2009). Of these 100, 17 were international players, and one played Division II basketball. This left 82 players who played Division I NCAA basketball. Since 34 of these players did not play in the NBA, there is no way to tell where they should have been drafted. But we can look at the predicted rankings and determine which draft round that translates to, and compare that to the predicted draft round of our formula.

Considering the proportion of international/non-Division I players in the prospects, the college players were grouped into top 25, 26-50, and 50+, with top 25 roughly representing 1st round pick, 26-50 roughly representing 2nd round pick, and 50+ being undrafted. With three unranked players having played in the NBA beginning that year, there are 85 players in this sample.

Visually, the best way to compare the predicted draft round with the actual draft round is with a 3x3 table. Along the top is the predicted draft rank, and along the side is the actual draft round. There are three tables comparing pre-draft rankings: M1 vs. Actual, M2 vs. Actual, and Big Board vs. Actual. Table 9 shows the three comparisons.

Table 9

3X3 Tables of M1, M2, and Pre-Draft Predicted Ranks vs. Actual Draft Round, 2009

M1		*Pred*			
		1-25	26-50	50+	
Actual	1st	13	7	4	24
	2nd	9	4	11	24
	Not Drafted	3	14	20	37
		25	25	35	85
M2		*Pred*			
		1-25	26-50	50+	
Actual	1st	11	4	9	24
	2nd	8	8	7	24
	Not Drafted	6	12	19	37
		25	25	35	85
Pre-Draft		*Pred*			
		1-25	26-50	50+	
Actual	1st	21	3	0	24
	2nd	4	11	9	24
	Not Drafted	0	11	26	37
		25	25	35	85

The Big Board does a very good job of predicting 1st round picks, 2nd round picks, and undrafted players. Since all other factors are taken into account (conference, physical measurements, intangibles, etc.), and the fact that teams only interview certain players, it is likely that these rankings will be fairly close to what actually happens. So comparing M1 to M2, M1 does a markedly better job in determining the future. There are far fewer extreme misses (i.e., a 1st round draft pick that is predicted to be undrafted by the model, or an undrafted player being predicted to go in the top 25, denoted in red in the table), and a higher rate of correct 1st round predictions.

Now for the players that did play in the NBA, but were not drafted, we can perform a similar analysis to the Spearman analysis, but we must now estimate the draft position of those players that were not drafted. Using the Big Board pre-draft rankings seemed like a good way to rank those players. There were 41 players taken in the 1st or 2nd round of the draft, so the highest ranked player on the Big Board not picked in the draft would receive a rank of 42, the next highest ranked player would get a rank of 43, and so on. In all, there were 50 players in 2009 that were either drafted and played, or signed as a free agent with a team. This ranking was used along with the M1 and M2 predicted rankings for the players. Each of these rankings was then compared against the “ideal” ranking of the players, based on their career WSP48. The differences were squared and summed, and the Spearman rank correlation coefficient was calculated for each comparison. Table 10 on the following page shows the results, as well as the results from the initial analysis for 2009 on page 49.

Table 10

Summary of Differences and Squared Differences between Ideal, Actual, and Predicted Rankings of WSP48, 2009 (College Statistics)

Year	Sum of $[ABS(d-i)]^2$ with Spearman	Sum of $[ABS(m1-i)]^2$ with Spearman	Sum of $[ABS(m2-i)]^2$ with Spearman
2009 (Top Prospects) n=50	15802 (0.241)	13837 (0.336)	17602 (0.155)
2009 (First round picks) n=24	1808 (0.214)	978 (0.575)	2158 (0.062)

Marked with yellow, M1 did the best job relative to the "ideal" position the players should have been chosen in, ranked by WSP48. M2 did worse than the actual draft order, which agrees with the results from the large sample analysis, where M1 is the superior choice in predicting draft order. The Spearman coefficient is nearly the same for the draft vs. ideal (0.241 and 0.214), while the M2 vs. ideal Spearman coefficient improves significantly (0.062 to 0.155). The M1 vs. ideal Spearman coefficient, while experiencing a large drop (0.575 to 0.336), is still the best method of the three for predicting success by WSP48. These results confirm the model M1 as an adequate, and potentially advanced, prediction tool to use when evaluating NBA prospects.

Chapter 7: Conclusion

Certain aspects of college players' careers have been highlighted by this paper as significant in predicting future success. Field goal percentage, blocks, and 1st team All-American were significant variables across all three initial analyses, indicating that talented centers have dominated the sport across these two decades. For each 1st team All-American a player makes in college, they can expect their career PER to be 1.32 higher than a player who did not make the 1st team. That player's career Win Shares will also be 14.123 higher, and Win Shares per 48 minutes will be .0181 higher for every 1st team they make. With the advent of the pace and space offense, and the emphasis placed on three point shooting, it will be interesting to see whether this trend continues.

The ability to predict the rest of a player's career based on their rookie season statistics has been proven by this study. Variables such as draft position, games, and personal fouls are significant across all three analyses. Draft position had been shown to be significant in predicting playing time in previous studies, and we have also shown that it is effective in predicting productivity on the court as well. It can be argued that games and personal fouls are closely tied, since a player who is playing more games will accumulate more personal fouls, just by being out on the court.

Rookie PER is also tied closely with Career PER, indicating that you can fairly well determine a player's performance for his career after one season. The same is true for rookie WS and rookie WSP48. However, there are limitations to these models.

In our sample, using college statistics (M1) to predict WSP48, there were eight players that the model predicted 20 or more spots away from their ideal ranking. The two that were predicted 20 spots higher were Bobby Hurley, and Ed O'Bannon. Both of these players were 1st team All-American. Hurley and O'Bannon were the NCAA Tournament Most Outstanding Player, which previous studies have shown will have an impact on draft pick. Both of these players were picked in the top 10 of their respective drafts. This shows that there are many outside factors that can influence a player's career PER. Hurley was involved in a car accident during his rookie season, which most likely shortened his career. O'Bannon had torn his ACL before his first year of college, and that led to him only being able to play two seasons in the NBA.

The six players that were undervalued, or picked 20 spots or more lower, were Dana Barros, Jayson Williams, Terrell Brandon, Steve Smith, Charlie Ward, and Steve Nash. Barros had a lower field goal percentage in college, and had low rebounds and blocks. Williams only played 13 games his final season in college, and that led to lower totals in general. Brandon had low rebound and block totals. Smith was a great scorer in college, but contributed very little in the way of steals and blocks. Ward only played 16 games and had an abysmal shooting percentage. Nash, the future Hall of Famer, actually had a low field goal percentage, low rebound and steal totals, and zero blocks the entire year.

Of all the players whose predictions using M1 were off by 20 or more, only one (Williams) was not a point guard or shooting guard. This was the reason to do individual regressions by position. This was not a perfect fit, however, as eight more players were predicted at least 20 spots away from their ideal ranking. Tate George, Michael Beasley, and Donte Greene were all predicted at least 20 spots too high. Beasley had a stellar college career, and was a universal high draft pick, but has never developed into a quality NBA player. George, a point guard, and Greene, a small forward, were artificially boosted by the simplicity of the formulas. George played for a team that was a #1 seed in the NCAA tournament, and that was the only variable used. Therefore, he had the lowest amount deducted from his projected WSP48 possible. Greene was a 6'11" small forward, which is incredibly tall for that position. Since height was significant for that position, he got a huge boost.

Scott Burrell, Chauncey Billups, Ryan Anderson, Steph Curry, and Paul George were all predicted at least 20 spots too low. Burrell, a small forward, was 6'7", and only played in 26 games. Both worked against his ranking. Billups, a point guard, played for a #9 seeded team, and was named 2nd team All-American, which is a negative coefficient for point guards. Anderson had a decent shooting percentage, but very low steals, which are the only two significant variables for power forwards, other than player of the year awards, which he did not win. Curry was penalized for not making the NCAA tournament, which was aided by the fact that he did not play in a major conference. This was the only factor that was accounted for in Curry's predicted rank. George was a very poor shooter in college, shooting only 42.4%. This, accompanied with his relatively low assists and blocks, led to his low ranking. These cases demonstrate the trouble with using only a few select statistics when predicting how a player will perform.

Of course, there are simply cases when both the real draft pick and the model value a player highly, but the player just does not meet his perceived potential. There are five players that both the model and the real draft had ranked a player at least 15 places too high in the draft: Sharone Wright, Michael Beasley, Joe Alexander, Jonny Flynn, and Evan Turner. With the exception of Wright, all players were drafted between 2008 and 2010, outside of our original sample. This could be attributed to the fact that they are not in the original sample, but Beasley and Turner won POY awards, and all had outstanding college careers. This is the variability that all owners and GM's will inevitably have to deal with during a draft.

The results from this study indicate that college statistics do provide a good model for predicting future NBA success, in terms of win shares per 48 minutes. The model used to predict the draft order did a significantly better job using WSP48 to that of the actual draft. These support my hypothesis that it is possible to adequately rank a player based solely on their college career. The results did not hold true for PER and win shares, however.

The final analysis of the Big Board Top 100 in 2009 for the draft demonstrated that M1 is the superior model when compared to M2, and also when compared to the draft for that particular year. M1 had a smaller difference between the predicted draft slots and the ideal draft slots, based on WSP48. Even with the regression equations being derived from only data on 1st round players, M1 was able to better predict how all NBA prospects should be drafted. Remarkably, all of the scouting and analysis done by NBA teams can be matched or improved upon by using M1. These results validate the entire premise of this paper, that purely quantitative methods can be used to predict which college players will have more successful NBA careers.

There are many different aspects that future studies could explore on this subject. This study completely ignored players from high school and international players. Since players like LeBron James, Kobe Bryant, and Kevin Garnett have all had Hall of Fame worthy careers, it would be interesting to see if there is any way to predict which high school players would have successful careers. It would also be worthwhile to get international data from other leagues, to see if those numbers can translate into the NBA. With the mix of success and failure of international players (future Hall of Famer Dirk Nowitzki vs. #1 overall pick and bust Andrea Bargnani), getting better methods of predicting a foreign players success would help shed some light on the topic.

There is also the NCAA PER variable, created by John Hollinger. This is a similar concept to PER in the NBA, but the data is only available since 2010. Since this does not give us a large enough time frame to adequately analyze how well NBA PER and NCAA PER are correlated, we will have to wait until enough data is available.

Drafting successfully in the NBA is one of the greatest keys to winning. Tim Duncan and Kobe Bryant have each won five championships, all with the one team they have played for their entire career. Even though Bryant did not get drafted by the Lakers directly, there was a trade in place to make sure he would end up in Los Angeles. The Lakers knew which player they wanted, and they struck gold with a pick in the middle of the first round. The ability to rank a player based on their college statistics is quite helpful, as well as the ability to re-evaluate a player after they have been in the league for a year. This paper has shown that this can be done with remarkable effectiveness.

References

1984 NBA draft. (n.d.). Retrieved November 12, 2014, from http://www.basketball-reference.com/draft/NBA_1984.html.

2007 NBA draft. (n.d.). Retrieved November 12, 2014, from http://www.basketball-reference.com/draft/NBA_2007.html.

A look at high school players in the NBA draft. (n.d.). Retrieved November 8, 2014, from https://www.rivals.com/content.asp?SID=1132&CID=356192.

Basketball-reference.com. (2014, June 30). *Basketball-Reference.com*. n.p.

Berri, D. J., Brook, S. L., & Schmidt, M. B. (2007). Does one simply need to score to score? *International Journal of Sport Finance, 2*(4), 190-205.

Berri, D. J., Schmidt, M. B., & Brook, S. L. (2006). *The wages of wins: Taking measure of the many myths in modern sport*. Stanford University Press.

Calculating PER. (n.d.). Retrieved November 4, 2014, from http://www.basketball-reference.com/about/per.html.

Calculating win shares. (2014, January 1). Retrieved November 4, 2014, from http://www.basketball-reference.com/about/ws.html.

Camerer, C. F., & Weber, R. A. (1999). The econometrics and behavioral economics of escalation of commitment: A re-examination of Staw and Hoang’s NBA data. *Journal of Economic Behavior & Organization, 39*(1), 59-82.

Chandler, G., & Stevens, G. (2012). An exploratory study of minor league baseball statistics. *Journal of Quantitative Analysis in Sports, 8*(4), n.p. De Gruyter, November 12, 2012.

Chang, J., & Zenilman, J. (2013, April 19). *A study of sabermetrics in major league baseball: The impact of Moneyball on free agent salaries*. Unpublished dissertation, Washington University in St. Louis.

Coates, D., & Oguntimein, B. (2008). The length and success of NBA careers: Does college production predict professional outcomes? *North American Association of Sports Economists*, n.p.

FAQ. (2014, January 1). Retrieved November 4, 2014, from http://wagesofwins.com/faq/.

Galletti, A. (2010, July 20). *Predictive stats, bad metrics & correlation in the NBA*. Retrieved November 4, 2014, from https://arturogalletti.wordpress.com/ 2010/07/20/predictive-stats-bad-metrics-correlation-in-the-nba/.

Golliver, B. (2012, December 13). Ex blazers exec: Portland wanted Bowie and Barkley over Jordan in 1984 draft. *SI.com*, n.p.

Groothuis, P. A., Hill, J. R., & Perri, T. (2007). The dilemma of choosing talent: Michael Jordans are hard to find. *Applied Economics, 41*(25), 3193-3198.

NBA.com-CBA principal deal points. (n.d.). Retrieved November 8, 2014, from http://www.nba.com/news/cba_summary_050804.html.

NBA.com: Evolution of the draft and lottery. (n.d.). Retrieved November 8, 2014, from http://www.nba.com/history/draft_evolution.htm.

NBA.com: Spencer Haywood bio. (n.d.). Retrieved November 8, 2014, from http://www.nba.com/history/players/haywood_bio.html.

Oliver, D. (2011). Basketball on paper: Rules and tools for performance analysis (1st ed.). Washington, DC: Potomac.

Player efficiency rating. (2014, October 24). Retrieved November 4, 2014, from http://en.wikipedia.org/wiki/Player_efficiency_rating.

Points produced definition-sporting charts. (n.d.). Retrieved November 12, 2014, from http://www.sportingcharts.com/dictionary/nba/points-produced.aspx.

Smith, A. (2009, June 24). *Aran Smith's 2009 Big Board: Top 100 Prospects*. Sports Phenoms, Inc.

Staw, B. M., & Hoang, H. (1995, September). Sunk costs in the NBA: Why draft order affects playing time and survival in professional basketball. *Administrative Science Quarterly, 40*(3), 474-494.

Table of contents. (n.d.). Retrieved November 2, 2014, from http://www.cbafaq.com/salarycap.htm.

Tymkovich, J. L. (2012, January). A study of minor league baseball prospects and their expected future value. Unpublished thesis, Claremont McKenna College.

Zminda, D. (2010). SABR. Henry Chadwick Award: Bill James. *Baseball Research Journal*.

Appendix A: Calculations for Win Shares

FOR OFFENSIVE WIN SHARES:

1. *Calculate Points Produced*

Points Produced = (FG Part + Ast Part + FT Part) * (1 – (TMOR/TMScPoss) * TMOR weight * (TMPlay%) + OR Part

Where

FGPart = 2 * (FGM + 0.5*(FG3M)) * (1 – 0.5*[(PTS-FTM)/2*FGA]*q_{ast}) estimation of how many points produced from 2-point and 3-point shots

And

Ast Part = 2 * [(TMFGM – FGM + 0.5*(TMFG3M – FG3M)/(TMFGM – FGM)] * 0.5 * [((TMPTS – TMFTM) – (PTS – FTM))/(2*(TMFGA-FGA))] * AST estimation of how many points one player's assists contributed to

And

FT Part = FTM total free throws made by player

TMOR = Team Offensive Rebounds

TMScPoss = Team Scoring Possessions

TMOR weight = [(1-TMOR%) * TMPlay%]/[(1-TMOR%)*TMPlay% + TMOR%*(1-TMPlay%)] estimation of weight applied to a team's offensive rebounds by how many offensive rebounds they get relative to how many possessions they have

TMPlay% = % the team scores on its possessions when offensive rebounds are considered a new possession

And

OR Part = OR * TMOR weight * TMPlay% * [TMPTS/(TMFGM + [1 – (1 – TMFT%)^2] * 0.4 * TMFTA)] accounts for offensive rebounds, weighted for how many opportunities the team has for offensive rebounds

And

Q_{ast} = (MIN/(TMMIN/5))*q_5 + [1-(MIN/(TMMIN/5))]*q_{12} estimation of how many points from your scoring was attributed to your teammate's assists.

Q_5 = 1.14*(TMAST-AST)/TMFGM a good estimation if the player is on the court for a large number of minutes

Q_{12} = [(TMAST/TMMIN) * MIN * 5 – AST]/[(TMFGM/TMMIN) * MIN * 5 – FGM] a good estimation if the player is on the court for a small number of minutes

2. *Calculate offensive possessions*

Possessions = Scoring Possessions + Missed FG Part + Missed FT Part + TOV

Where

Scoring Possessions = (FG Part + AST Part + FT Part) * (1 – TMOR/TMScPoss * TMOR weight * TMPlay%) + OR Part

And
Missed FG Part = (FGA-FGM) * (1 – 1.07*TMOR%)
And
Missed FT Part = $(1 – FT\%)^2$ * 0.4 * FTA

3. *Calculate marginal offense*

Marginal Offense = Points Produced – 0.92*(league points per possession)*(offensive possessions)
Points produced and offensive possessions were calculated in steps 1 and 2, respectively. League points per possession is just a league wide average of points scored per possession. The 0.92 multiplier is an estimation from a Pythagorean expectation equation, namely:

$(1 - z)^{14} / ((1 - z)^{14} + (1 + z)^{14})) = 0.10$

The creator of the formula assumed that a team of replacement players would win about 10% of their games. The exponent of 14 is a rounded version of the exponent used in the NBA's Pythagorean expectation equation:

Team Wins = points $for^{13.91}$ / points $for^{13.91}$ + points $against^{13.91}$

This was created by Daryl Morey, owner of the Houston Rockets, to predict how many wins a team would have based on their points scored and points allowed. The exponent was estimated at 13.91 because it did a better job of predicting actual wins.
So solving for z in the first equation, the author came up with 0.92, which shrinks the number of expected points a team will score, ensuring that the marginal offense for a team is non-negative. This is important because the players most responsible for team success will get the largest share of marginal offense. If the marginal offense is negative, they are punished by having more of the negative marginal offense.

4. *Calculate marginal points per win*

Marginal points per win = 0.32 * (league points per game) * ((team pace)/(league pace)).
League points per game, team pace, and league pace are all recorded and calculated by the NBA. 0.32, to the best of my knowledge, is a multiplier to account for statistical error. Basically, marginal points per win is a measurement of one team's production in relation to the league average. This is on a TEAM level, not an INDIVIDUAL level.

5. *Calculate Offensive Win Shares*

Offensive Win Shares = (Marginal Offense)/(Marginal Points Per Win)

FOR DEFENSIVE WIN SHARES:

1. *Calculate the Defensive Rating*

DRtg = TMDRtg + 0.2 * [100 * DPtsPerScPoss * (1 – Stop%) – TMDRtg]
Where
TMDRtg = # of points a team gives up per 100 possessions

DPtsPerScPoss = # of points scored per scoring possession by the team's opponents
Stop% = (Stops*TMMIN)/(TMPOSS*MIN)
Defensive rating is how many points a player gives up per 100 possessions

2. *Calculate Marginal Defense*

Marginal Defense = (Player minutes played / team minutes played) * (team defensive possessions) * (1.08 * (league points per possession) – ((Defensive Rating)/100))
The 1.08 multiplier was created along the same logic as #3 in the Offensive Win Shares calculation.

3. *Calculate marginal points per win*

Marginal points per win = 0.32 * (league points per game) * ((team pace)/(league pace))
This is the same as in the Offensive Win Shares

4. *Calculate Defensive Win Shares*

Defensive Win Shares = (Marginal Defense)/(Marginal Points Per Win)

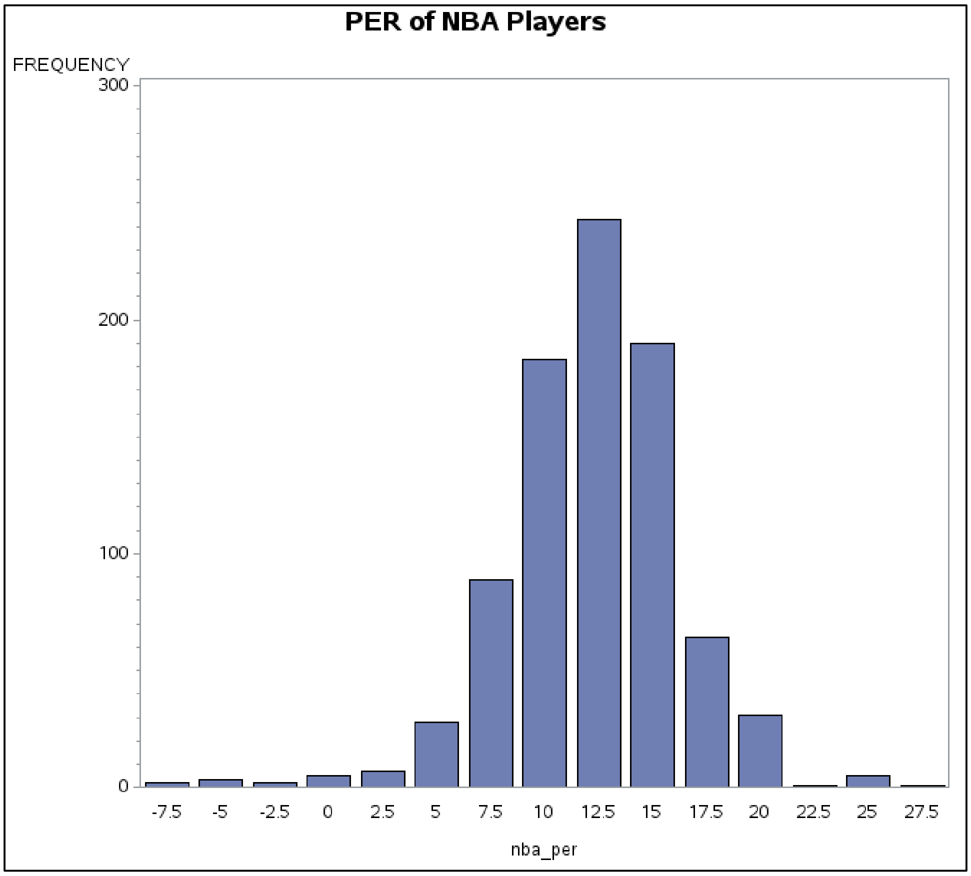
FOR WIN SHARES:
Win Shares = Offensive Win Shares + Defensive Win Shares

CALCULATING PER

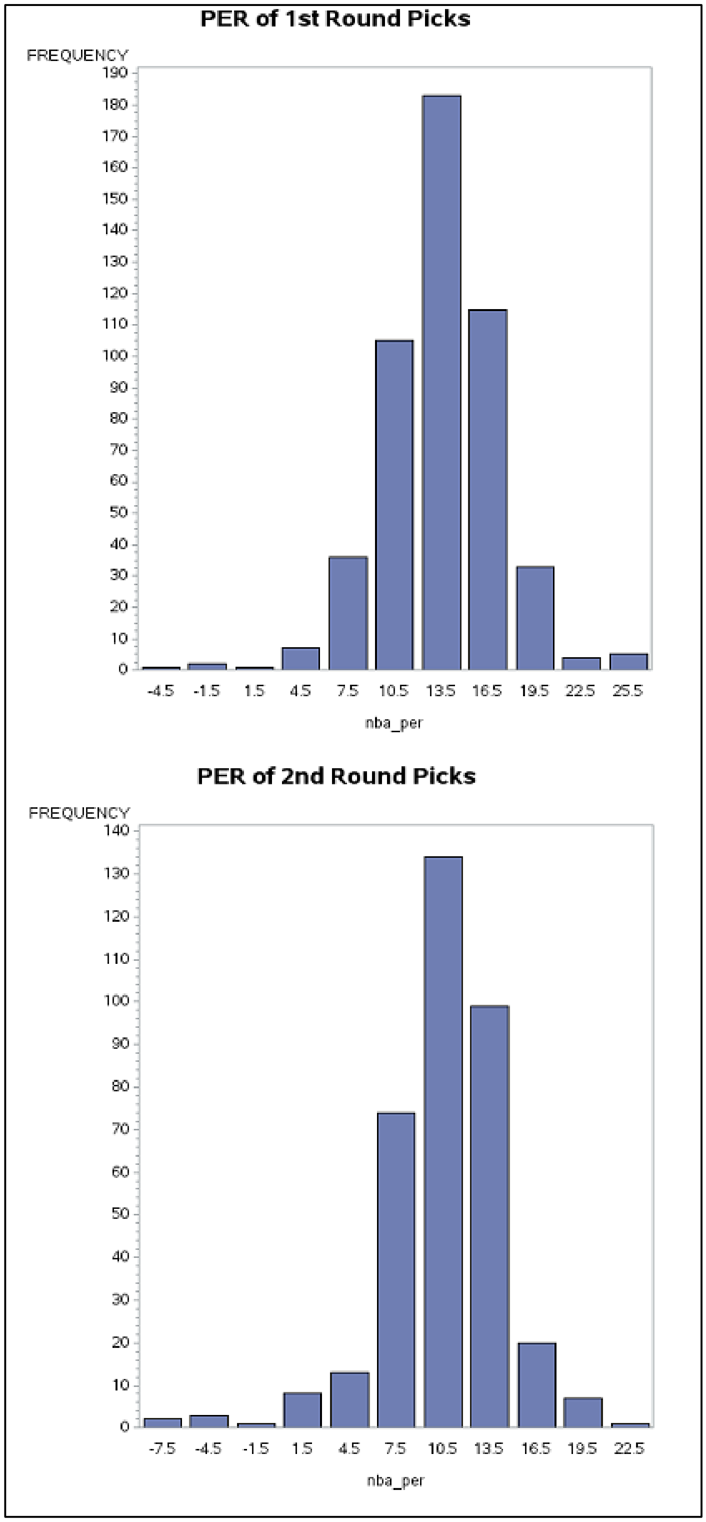
Step 1: uPER = (1 / MP) *
[3P
+ (2/3) * AST
+ (2 - factor * (team_AST / team_FG)) * FG
+ (FT *0.5 * (1 + (1 - (team_AST / team_FG)) + (2/3) * (team_AST / team_FG)))
- VOP * TOV
- VOP * DRB% * (FGA - FG)
- VOP * 0.44 * (0.44 + (0.56 * DRB%)) * (FTA - FT)
+ VOP * (1 - DRB%) * (TRB - ORB)
+ VOP * DRB% * ORB
+ VOP * STL
+ VOP * DRB% * BLK
- PF * ((lg_FT / lg_PF) - 0.44 * (lg_FTA / lg_PF) * VOP)]
Where
factor = (2 / 3) - (0.5 * (lg_AST / lg_FG)) / (2 * (lg_FG / lg_FT)) (a normalizing factor, I'm assuming)
VOP = lg_PTS / (lg_FGA - lg_ORB + lg_TOV + 0.44 * lg_FTA) (Value of Possession. An estimation of how many points per possession the league has scored)
DRB% = (lg_TRB - lg_ORB) / lg_TRB (Rate of defensive rebounds)

Step 2: pace adjustment = lg_Pace / team_Pace
Step 3: aPER = (pace adjustment) * uPER
Step 4: PER = aPER * (15 / lg_aPER)

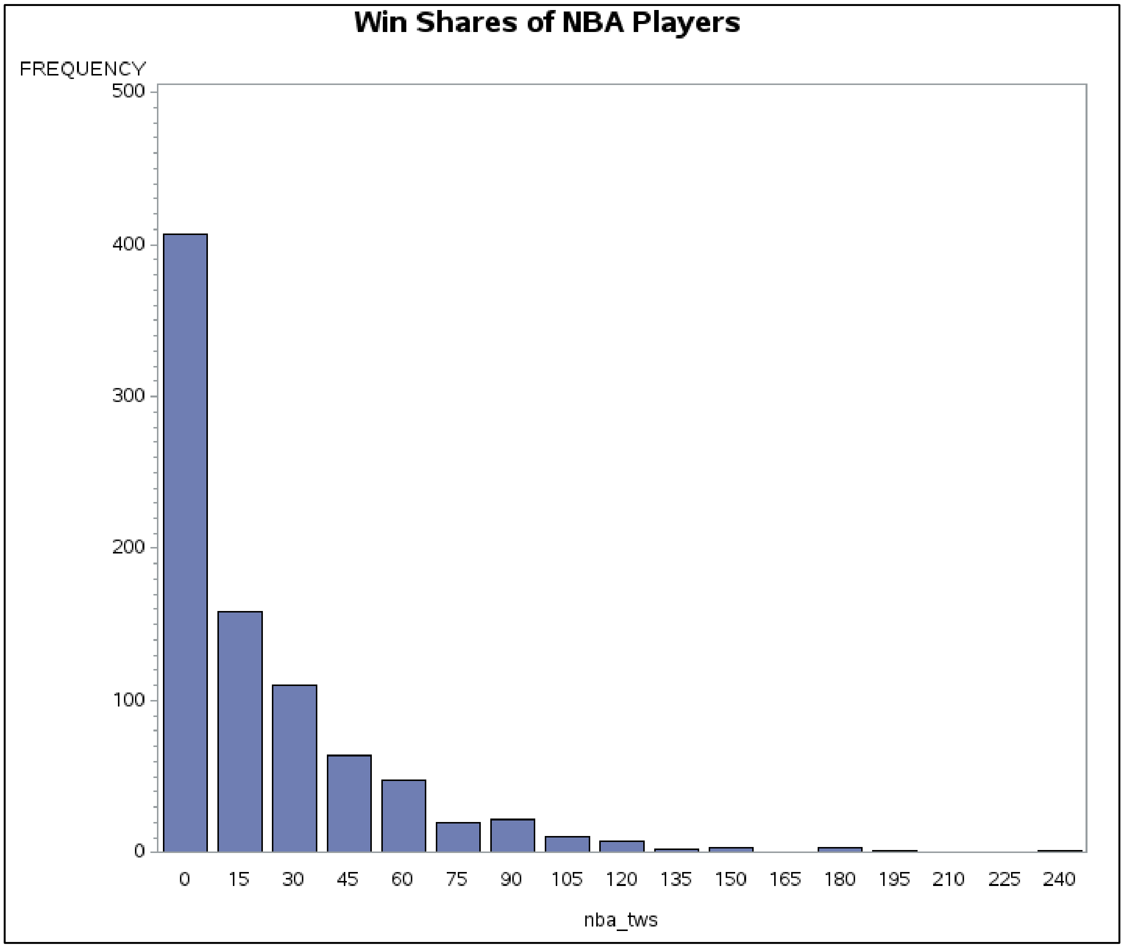
Appendix B: Statistics for PER, Win Shares, and Win Shares per 48 Minutes

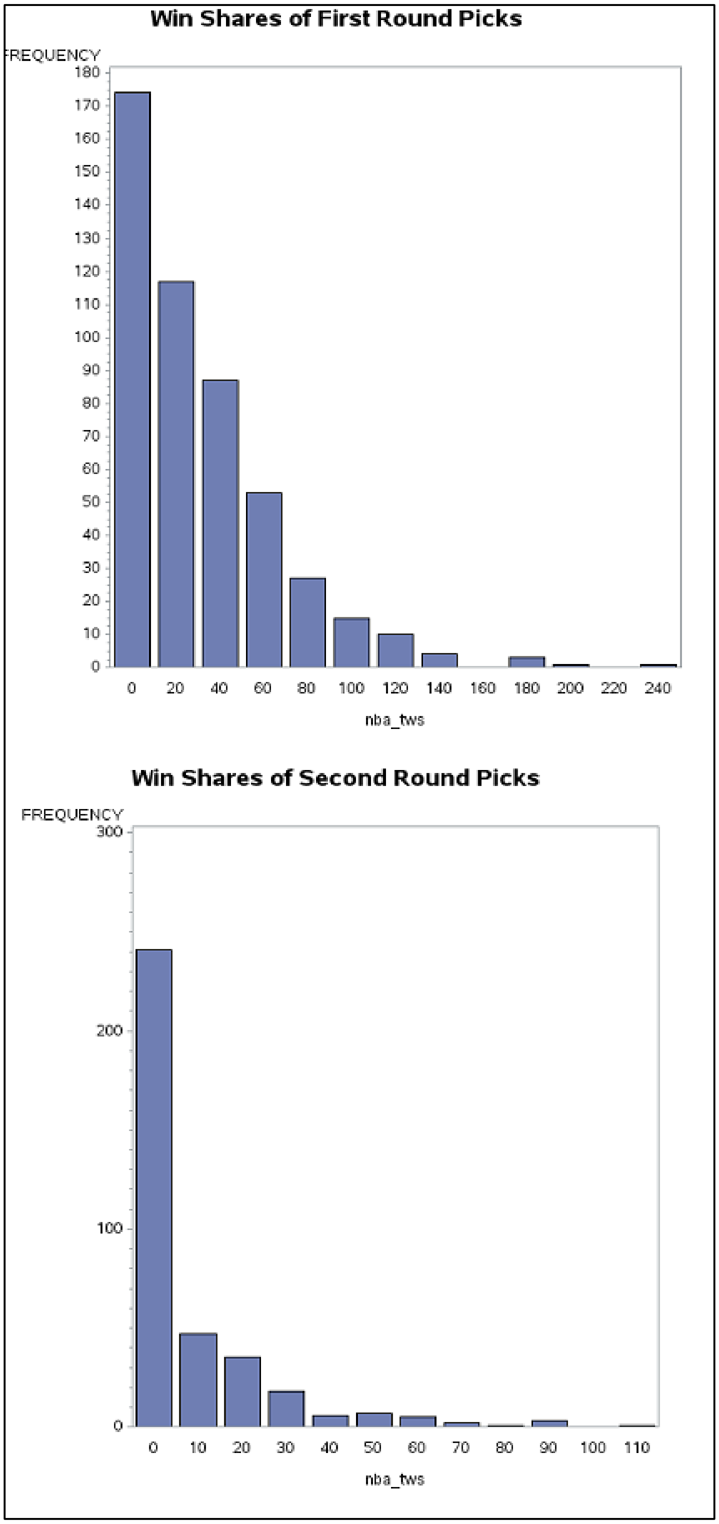


Career PER of 1st and 2nd Round Picks with NBA and NCAA Careers in Our Sample (1985-2005)

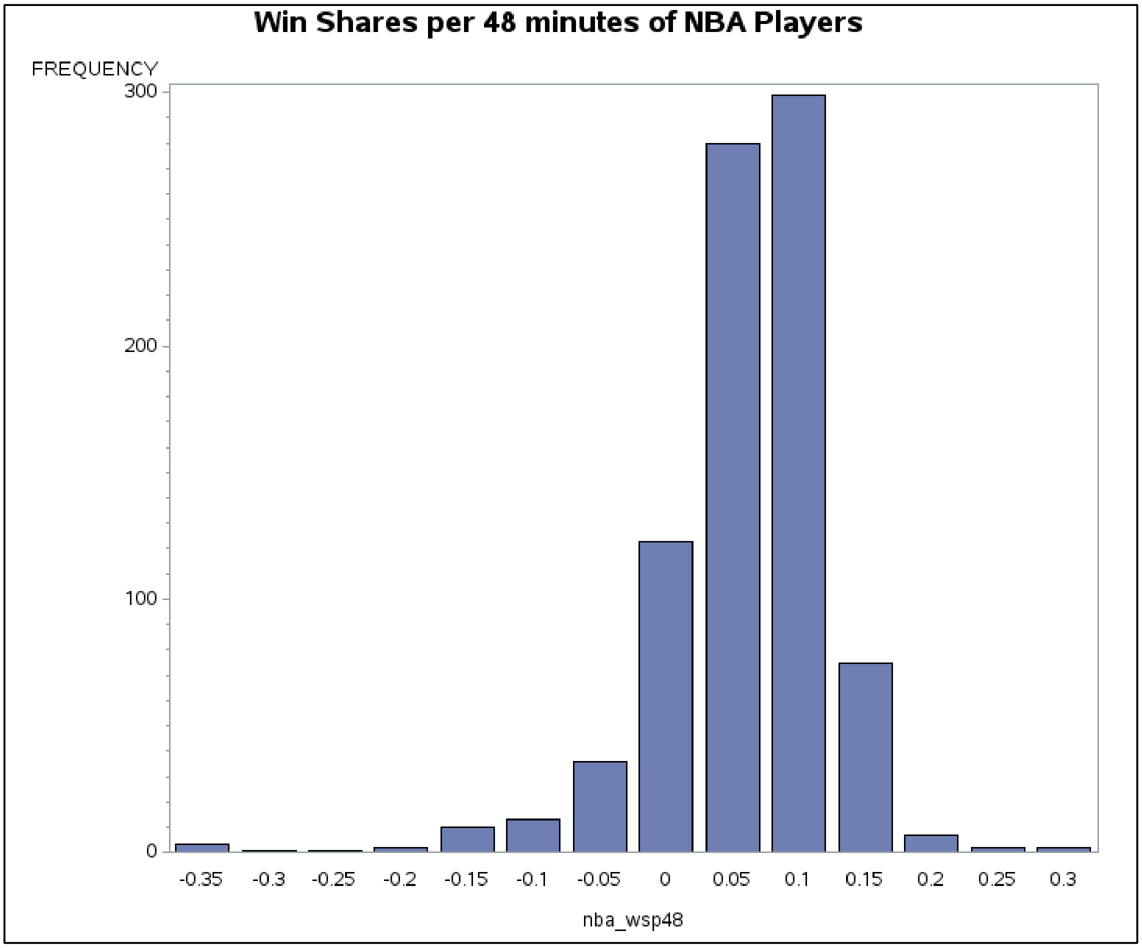


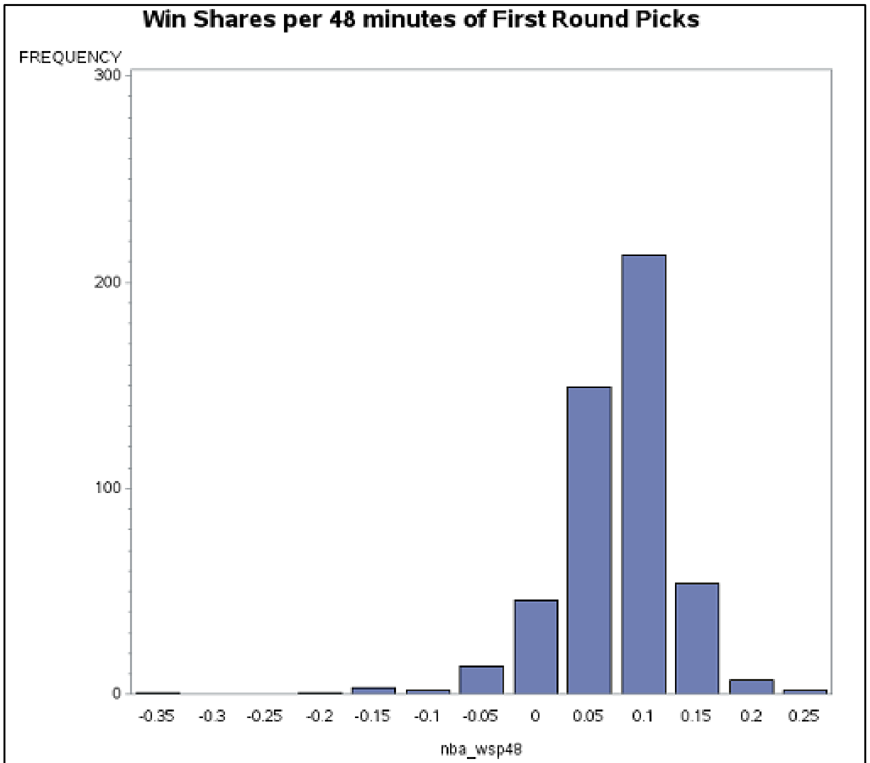
Career WS of 1st and 2nd Round Picks with NBA and NCAA Careers in Our Sample (1985-2005)

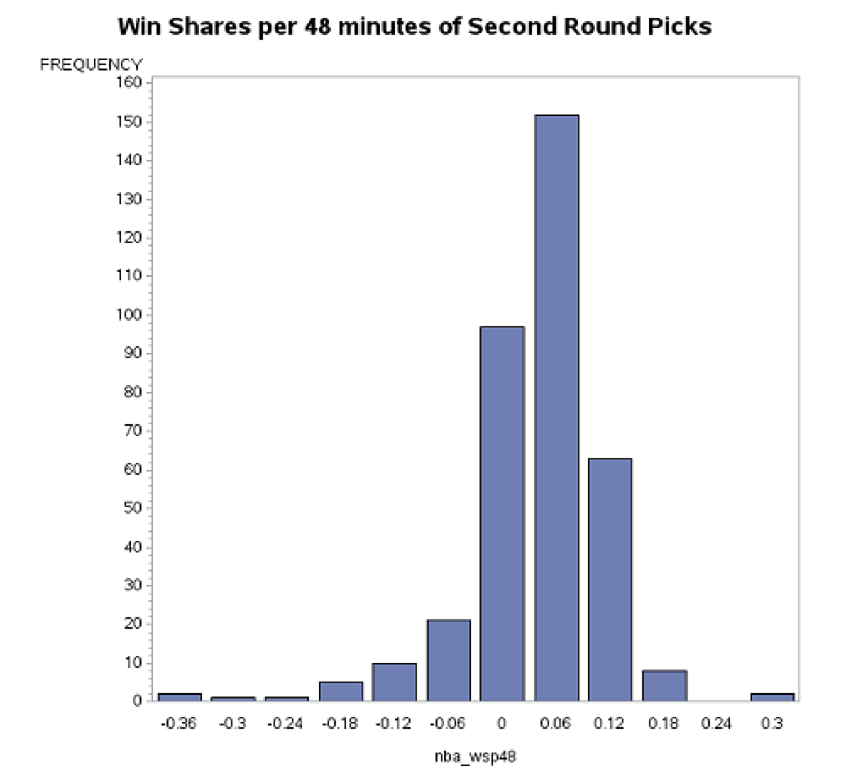

Win Shares of NBA Players
FREQUENCY
500
400
300
200
100
0
0
15
30
45
60
75
90
105
120
135
150
165
180
195
210
225
240
nba_tws



Career WSP48 of 1st and 2nd Round Picks with NBA and NCAA Careers in Our Sample (1985-2005)


Win Shares per 48 minutes of NBA Players
FREQUENCY
300
200
100
0
-0.35
-0.3
-0.25
-0.2
-0.15
-0.1
-0.05
0
0.05
0.1
0.15
0.2
0.25
0.3
nba_wsp48


Win Shares per 48 minutes of First Round Picks
FREQUENCY
300
200
100
0
-0.35
-0.3
-0.25
-0.2
-0.15
-0.1
-0.05
0
0.05
0.1
0.15
0.2
0.25
nba_wsp48


Win Shares per 48 minutes of Second Round Picks
FREQUENCY
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-0.36
-0.3
-0.24
-0.18
-0.12
-0.06
0
0.06
0.12
0.18
0.24
0.3
nba_wsp48

Appendix C: 1st and 2nd Round Regression Analysis

Table A

Regression of College Variables against NBA PER, Win Shares, and Win Shares per 48 Minutes, 1st Round Only (N = 477)

Variable	**PER**	**Win Shares**	**Win Shares per 48 Minutes**
Intercept	4.375(1.694)	-43.394(16.177)	-0.0799(0.0292)
Field Goals Made			-0.0004***(0.0002)
Field Goal Percentage	8.479***(3.122)	62.725**(29.778)	0.2238***(0.0536)
Points			0.0001*(0.0001)
Rebounds			0.0001***(0)
Assists		0.122***(0.037)	0.0001**(0.0001)
Steals	0.03***(0.009)		
Blocks	0.017***(0.006)	0.109*(0.056)	0.0002**(0.0001)
Rebounds Per Game	0.158*(0.081)	1.422*(0.78)	
Assists Per Game	0.462***(0.124)		
Steals Per Game		10.334***(2.754)	0.015***(0.0047)
1st Team All-American	1.32***(0.372)	14.123***(3.702)	0.0181***(0.0054)
AP Player of the Year	1.484*(0.856)	35.949***(10.607)	
NABC Defensive Player of the Year		9.495*(5.691)	
NABC Player of the Year		-22.89**(9.766)	
USBWA Freshman of the Year	1.784*(0.939)		
Summary Statistics			
R^2	.207	.213	.1149

Standard errors are given in parentheses next to the coefficients. The coefficients are significant at the *10%, **5%, and ***1% levels.

Table B

Regression of College Variables against NBA PER, Win Shares, and Win Shares per 48 Minutes, 2nd Round Only (N = 336)

Variable	**PER**	**Win Shares**	**Win Shares per 48 Minutes**
Intercept	8.827(1.703)	-23.912(11.602)	-0.081(0.086)
Field Goals Made	0.015(0.009)		
Field Goal Percentage		48.696**(20.717)	0.28(0.171)
Steals		0.214***(0.061)	
2nd Team All-American		8.18(5.445)	
3rd Team All-American		12.851***(4.594)	
Summary Statistics			
R^2	.011	.109	.048

Standard errors are given in parentheses next to the coefficients. The coefficients are significant at the *10%, **5%, and ***1% levels.

Appendix D: 1st Round by Position Regression Analysis

Table C

Regression of College Variables against NBA PER, Win Shares, and Win Shares per 48 Minutes, Point Guards, 1st Round Only (N = 84)

Variable	**PER**	**Win Shares**	**Win Shares per 48 Minutes**
Intercept	2.954(4.803)	-28.721(18.802)	0.1079(0.0144)
Field Goal Percentage	16.246(10.13)		
Rebounds		0.207**(0.101)	
Steals	0.042**(0.021)		
Assists Per Game		7.113***(2.273)	
Tournament Seed			-0.002(0.0013)
2nd Team All-American	-2.402***(0.853)		-0.048***(0.0145)
3rd Team All-American	-1.834*(1.047)		-0.0416**(0.0184)
Major Conference	1.984*(1.039)		
Summary Statistics			
R^2	.192	.139	.142

Standard errors are given in parentheses next to the coefficients. The coefficients are significant at the *10%, **5%, and ***1% levels.

Table D

Regression of college variables against NBA PER, Win Shares, and Win Shares per 48 minutes, Shooting Guards, 1st round only (N = 99)

Variable	**PER**	**Win Shares**	**Win Shares per 48 Minutes**
Intercept	2.83(3.813)	-2.996(9.663)	-0.082(0.062)
Field Goal Percentage	13.023*(7.744)		0.205(0.125)
Blocks	0.065**(0.027)	0.537**(0.268)	
Assists Per Game	1.007***(0.27)	6.879**(2.777)	0.013**(0.004)
Blocks Per Game			0.025*(0.014)
AP POY		17.237**(8.587)	
NIT MVP		45.632**(18.903)	
Summary Statistics			
R^2	.222	.196	.157

Standard errors are given in parentheses next to the coefficients. The coefficients are significant at the *10%, **5%, and ***1% levels.

Table E

Regression of College Variables against NBA PER, Win Shares, and Win Shares per 48 Minutes, Small Forwards, 1st Round Only (N = 95)

Variable	**PER**	**Win Shares**	**Win Shares per 48 Minutes**
Intercept	8.719(1.7952)	32.182(12.018)	-1.2488(0.4373)
Height			0.0089*(0.0046)
Games			0.0164***(0.006)
Free Throws Made		-0.157**(0.079)	
Rebounds	0.0144***(0.0046)		-0.0019**(0.0009)
Assists	0.0188**(0.009)	1.002***(0.317)	0.0003*(0.0002)
Rebounds Per Game			0.0689**(0.0281)
Assists Per Game		-25.166**(10.14)	
Team Losses	-0.1549*(0.0879)		
1st Team All-American		13.3(8.042)	
NABC DPOY		10.918(9.439)	
Major Conference	1.2719*(0.713)		0.0207(0.013)
Summary Statistics			
R^2	.204	.219	.214

Standard errors are given in parentheses next to the coefficients. The coefficients are significant at the *10%, **5%, and ***1% levels.

Table F

Regression of College Variables against NBA PER, Win Shares, and Win Shares per 48 Minutes, Power Forwards, 1st Round Only (N = 113)

Variable	**PER**	**Win Shares**	**Win Shares per 48 Minutes**
Intercept	27.434(5.996)	105.968(42.338)	-0.0522(0.0455)
Field Goals Attempted		0.021**(0.009)	
Field Goal Percentage			0.2064**(0.081)
Free Throws Percentage	-8.011**(3.881)	-76.076**(36.984)	
Rebounds	0.012**(0.005)		
Steals			0.0006**(0.0003)
Rebounds Per Game		2.747(1.8)	
Steals Per Game		11.102(7.198)	
Team Losses	-0.556***(0.207)	-4.31***(1.426)	
Conference Win %	-9.873**(4.646)	-2.865(1.773)	
1st Team All-American	1.169*(0.63)		
AP POY	2.484*(1.391)	28.381**(14.285)	0.0256(0.0206)
NABC DPOY		31.121***(11.62)	0.0281*(0.0166)
Summary Statistics			
R^2	.259	.385	.171

Standard errors are given in parentheses next to the coefficients. The coefficients are significant at the *10%, **5%, and ***1% levels.

Table G

Regression of College Variables against NBA PER, Win Shares, and Win Shares per 48 Minutes, Centers, 1st Round Only (N = 96)

Variable	**PER**	**Win Shares**	**Win Shares per 48 Minutes**
Intercept	19.904(4.308)	41.038(39.713)	0.1837(0.0623)
Weight	-0.045***(0.017)	-0.389***(0.134)	-0.0005**(0.0002)
Field Goal Percentage		76.996(49.421)	
Free Throws Made	0.012*(0.007)		
Blocks	0.031***(0.009)	0.198***(0.069)	
Rebounds Per Game		2.12(1.339)	
Blocks Per Game			0.0097**(0.0038)
1st Team All-American	1.915**(0.877)		0.0238*(0.0127)
2nd Team All-American		12.115(8.339)	
AP POY	1.051(2.234)	159.143***(24.725)	0.0073(0.0325)
Naismith Award		-152.633***(32.813)	
Sporting News POY		80.655***(25.922)	
NIT MVP	4.751(3.022)		
Summary Statistics			
R^2	.504	.645	.344

Standard errors are given in parentheses next to the coefficients. The coefficients are significant at the *10%, **5%, and ***1% levels.

Appendix E: Regression Results

Table H

Regression of Rookie Year Variables against PER, WS, and WSP48, 1st Round Only (N = 425)

Variable	**PER**	**Win Shares**	**Win Shares Per 48 Minutes**
Intercept	6.207(2.11)	9.248(6.388)	-0.0042(0.0209)
Weight	-0.012*(0.006)		
Draft Position	-0.049***(0.019)	-0.732***(0.208)	-0.0009***(0.0003)
Games	0.08***(0.015)	0.253**(0.108)	0.0013***(0.0002)
Free Throws Made		-0.047(0.03)	
Free Throw Percentage	-2.082*(1.216)		-0.0462**(0.0189)
Steals	-0.032***(0.008)		
Turnovers	0.015(0.01)		-0.0001(0.0002)
Personal Fouls	-0.04***(0.008)	-0.137***(0.034)	-0.0005***(0.0001)
Steals per game			-0.0178*(0.0096)
Turnovers per game	0.131(0.661)	13.967***(2.679)	0.0175(0.0112)
Personal Fouls per game	2.12***(0.56)		0.0309***(0.0098)
Points per 36 minutes	-0.041(0.048)		
Steals per 36 minutes	1.023***(0.337)		0.0153***(0.0059)
PER	0.514***(0.056)		
Offensive Win Shares	0.201*(0.115)		
Defensive Win Shares	0.633***(0.177)	4.346*(2.326)	
Win Shares		7.538***(1.371)	0.0038**(0.0016)
Win Shares per 48 Minutes		49.862(29.666)	0.4723***(0.0457)
All-Rookie 1st Team	0.646(0.409)		
Summary Statistics			
R^2	.634	.487	.559

Standard errors are given in parentheses next to the coefficients. The coefficients are significant at the *10%, **5%, and ***1% levels.

Table I

Regression of Rookie Year Variables against PER, WS, and WSP48, 1st Round Only, Point Guards (N = 84)

Variable	**PER**	**Win Shares**	**Win Shares Per 48 Minutes**
Intercept	8.47(2.735)	43.95(7.987)	0.03250(0.01017)
Draft Position		-1.428***(0.419)	
Games	0.038**(0.018)		
Field Goals Made			0.00011**(0.00005)
Field Goal Percentage	-15.899*(9.246)		
PER	0.78***(0.141)		
Win Shares		7.84***(1.571)	
Win Shares per 48 Minutes			0.53177***(0.08444)
Summary Statistics			
R^2	.532	.405	.472

Table J

Regression of Rookie Year Variables against PER, WS, and WSP48, 1st Round Only, Shooting Guards (N = 99)

Variable	**PER**	**Win Shares**	**Win Shares Per 48 Minutes**
Intercept	8.053(1.265)	21.725(10.401)	-0.01028(0.075)
Weight			-0.00045(0.0003)
Draft Position		-1.138***(0.423)	-0.00118**(0.0006)
Field Goal Percentage			0.45003***(0.1015)
Three Point Field Goals			0.00031***(0.0001)
Minutes per game	0.056*(0.032)		
Points per 36 minutes	-0.142(0.094)		-0.00346***(0.0012)
Rebounds per 36 minutes	-0.484**(0.23)		
Assists per 36 minutes		4.491**(1.89)	0.00317(0.0026)
Blocks per 36 minutes			0.03601***(0.0126)
PER	0.678***(0.107)		
Defensive Win Shares		-8.041*(4.631)	
Win Shares		11.028***(2.013)	
Win Shares per 48 Minutes			0.26837***(0.0962)
Rookie of the Year	2.519*(1.447)		
Summary Statistics			
R^2	.588	.455	.636

Table K

Regression of Rookie Year Variables against PER, WS, and WSP48, 1st Round Only, Small Forwards (N = 95)

Variable	**PER**	**Win Shares**	**Win Shares Per 48 Minutes**
Intercept	4.509(0.921)	-11.099(15.707)	0.04688(0.05)
Draft Position		-0.787**(0.394)	-0.00181***(0.0006)
Games			0.00085***(0.0003)
Field Goals Made		-0.137***(0.038)	-0.00022***(0.0001)
Free Throw Percentage			-0.09027**(0.0426)
Minutes per game	0.079***(0.024)	2.282***(0.601)	0.00248***(0.0008)
Blocks per 36 minutes	1.016*(0.588)		
Turnovers per 36 minutes			-0.00872*(0.0049)
PER	0.538***(0.079)	2.347**(1.055)	0.00625**(0.0024)
Win Shares		5.478***(1.924)	
Win Shares per 48 Minutes			0.1809(0.129)
Summary Statistics			
R^2	.555	.501	.618

Table L

Regression of Rookie Year Variables against PER, WS, and WSP48, 1st Round Only, Power Forwards (N = 113)

Variable	**PER**	**Win Shares**	**Win Shares Per 48 Minutes**
Intercept	44.045(15.764)	2.102(30.465)	0.02734(0.0214)
Height	-0.487**(0.192)		
Games		-0.278(0.19)	
Field Goal Percentage		131.786**(57.876)	
Three Pointers Made	0.033***(0.012)		
Free Throw Percentage		-69.255**(33.602)	
Minutes per game		0.904**(0.447)	0.00133***(0.0004)
Points per 36 minutes	-0.418***(0.144)		-0.00322**(0.0014)
Rebounds per 36 minutes			-0.00295(0.0018)
Assists per 36 minutes	-0.52*(0.288)		
Steals per 36 minutes			0.01466*(0.0076)
Blocks per 36 minutes		-10.144**(4.023)	
Turnovers per 36 minutes	1.98***(0.465)		0.01534***(0.0058)
PER	0.845***(0.119)		
Defensive Win Shares	0.546**(0.247)	18.175***(3.939)	
Offensive Win Shares			-0.01045***(0.0036)
Win Shares per 48 Minutes			0.90476***(0.0944)
# of Rookie of the Month		6.49**(3.06)	0.00852**(0.004)
Summary Statistics			
R^2	.668	.621	.659

Table M

Regression of Rookie Year Variables against PER, WS, and WSP48, 1st Round Only, Centers (N = 96)

Variable	**PER**	**Win Shares**	**Win Shares Per 48 Minutes**
Intercept	7.469(1.958)	39.664(29.9)	0.06004(0.0453)
Weight		-0.188(0.115)	-0.00037**(0.0002)
Games		-0.381***(0.137)	
Field Goal Percentage	-10.431**(4.505)		
Points per 36 minutes			-0.00633***(0.0018)
Assists per 36 minutes			-0.00927*(0.0055)
Blocks per 36 minutes	0.353(0.256)		
Turnovers per 36 minutes			0.01676**(0.0064)
PER	0.817***(0.116)	1.957***(0.612)	0.01189***(0.0016)
Defensive Win Shares	0.584**(0.288)	25.23***(2.653)	0.01267***(0.0037)
Win Shares per 48 Minutes	-18.179***(5.983)		
All Rookie 1st Team	2.173**(0.987)		
All Rookie 2nd Team	-1.488**(0.718)	-18.456**(7.287)	-0.02213**(0.011)
Summary Statistics			
R^2	.758	.824	.802