Generating Relative Draft Value in the NBA Draft and Predicting Success from College Basketball

A Major Qualifying Project Report:

submitted to the faculty of the

**WORCESTER POLYTECHNIC INSTITUTE**

in partial fulfillment of the requirements for the

degree of Bachelor of Science

by

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*Michael Krebs*

­\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*Jake Scheide*

Date:

Approved: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Professor Craig Wills, Major Advisor

# Abstract

# Acknowledgements

# Table of Contents

# Table of Figures

# Executive Summary

# 1. Introduction

This project’s goals are threefold: firstly, to evaluate the accuracy of current and developed player performance metrics for basketball players in the National Basketball Association (NBA). Secondly, to identify the relative draft value for the sixty picks in the NBA Draft. Finally, we present a model which predicts the draft position and performance for an individual based on their National Collegiate Athletic Association (NCAA) Division I statistics.

To achieve our first goal, we first select four of the most popular player performance metrics in the NBA, and analyze what sort of players and careers they favor. Additionally, we create two custom metrics which provide contrast to the established statistics and generate additional insight into why certain metrics are used.

Taking these insights into account, we then incorporate draft data to approximate how valuable each selection in the NBA Draft is, both relative to a player and another draft pick. We create an National Football League (NFL)-style draft value chart, and plot each of our six metrics and observe how talent changes with draft position comparably to salary.

For our third goal, we use ­\_\_\_\_\_\_\_\_\_ to construct a model which predicts the expected draft position for a player in a draft year, as well as their projected success in the NBA.

Through this process, we will be able to contextualize the plethora of player performance statistics, examine the value teams assign to draft picks compared to our statistical analysis, and

# 2. Background

Basketball on Paper – Dean Oliver

Assessing the Relative Value of Draft Position in the NBA Draft - By Aaron Barzilai, Ph.D.

# 3. Methods

To ensure we were extracting a valid performance metric for NBA players, we needed to look at more than just one number. Fortunately, many statisticians have tried to represent the worth of a

Table 1: Metrics used

|  |
| --- |
| Metrics Used |
| Player Efficiency Rating  Win Shares  Value Over Replacement Player  Fantasy Points  Basic Percentile  Advanced Percentile |

player in a single figure. The most common metric used when comparing NBA players is John Hollinger’s Player Efficiency Rating (PER), which is where we began our research. PER is calculated primarily using offensive box score statistics, a common problem identified by Dean Oliver, widely regarded as the pioneer of advanced basketball analytics. PER is normalized to both minutes played and pace, which is an important distinction we will analyze further later. A statistic which does not normalize to minutes played, but instead contribution to wins, is Win Shares (WS). Win Shares in basketball is derived from Bill James’ work in baseball, and follows a similar format. It attempts to reflect how many wins a particular player contributed to his team for a season, whilst taking into account pace of play. When evaluating these metrics’ ‘true’ accuracy, we will investigate how these normalizations warp the rankings. Value Over Replacement Player (VORP) was our third metric analyzed in our project. VORP estimates overall contribution to a team, by first creating an adjusted plus-minus statistic, then normalizing to minutes played and the value a theoretical minimum-salary player would bring in replacement of that player. An important note is that VORP was created to scale linearly with salary, so that a player with twice the VORP should be worth twice the salary. The fourth and final non-feature engineered statistic we used was Fantasy Points (FP). Fantasy Points is a simple metric calculated by weighting each of the basic box score statistics (Points, Rebounds, Assists, Blocks, and Steals) as well as turnovers and producing a sum. We chose to use FP because it provides a different angle to player evaluation- one which fans will appreciate. FP is not normalized to minutes played nor pace.

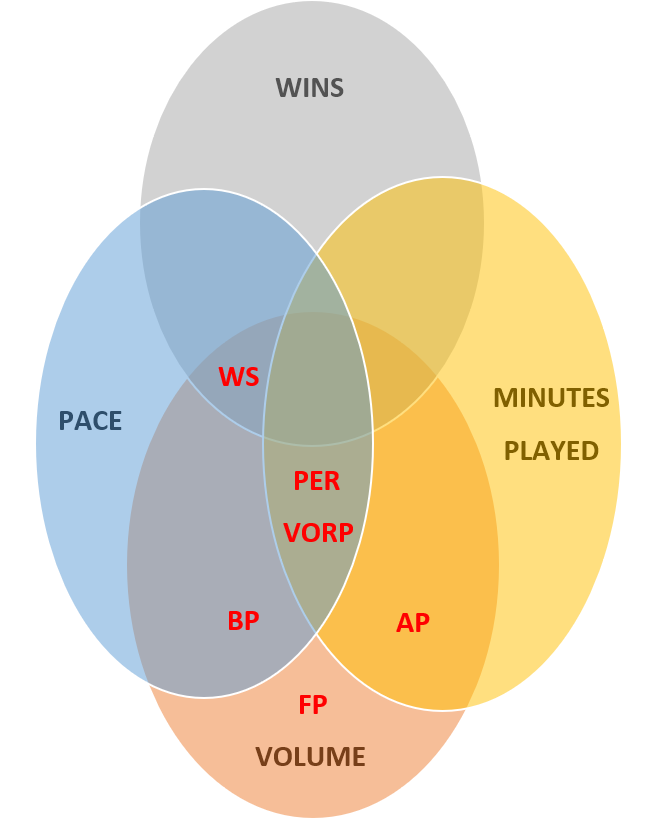
After researching existing metrics and developing an understanding of what already exists, we set out to create our own metric that measures a different variant of player performance. Basic Percentile (BP) ranks a player on their basic box score stats relative to the rest of the league. This means a player who leads the league in Points, Rebounds, Assists, Steals and Blocks would receive a score of 5 for a season. A player first in all categories except third in steals would receive a 7 (1+1+1+3+1). This naturally normalizes to league pace because it takes into account the raw volume from every other team that year. Advanced Percentile (AP) follows a similar train of thought, but it instead looks at True Shooting Percentage (TS%), Rebound Percentage, Assist Percentage, Steal Percentage, and Block Percentage. The percentage metrics evaluate the proportion of total box score stats available a player amasses while they are on the court. Therefore, we can say that AP naturally normalizes to minutes played.

Figure : Classification of Metrics by Normalization

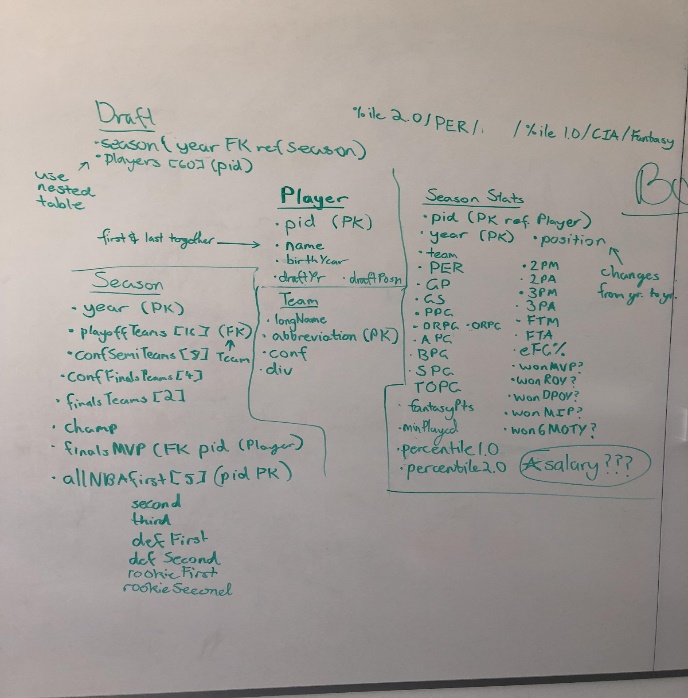
Our first drafted Entity Relationship Diagram is shown to the right. In the end, we modified this slightly to have a Master Season sheet for every year with all player season stats, a Draft sheet with the players drafted that season, and a Master Players sheet with all players and their IDs. With a clear goal with how we were going to structure our data we then set out to collect the data. Initially, we were going to reach out to a 3rd party and obtain a licensing fee for their data. As students, we were not able to access their data, and so we looked to online resources. Using Python 3.7 we incorporated the BeautifulSoup4 library to web crawl the Basketball-Reference.com website. From this website we were able to obtain all of the single season metrics for every player, all single season awards, the draft selections for every player who played in 1990 to 2018 and utilized the player unique id from basketball reference as our own primary key.

Figure 2: Initial Entity Relationship Diagram (Convert to Lucid Chart)

In order to obtain all of this data we searched through the <table> tags on the webpage, used the <th> tags as column headers, used the <tr> tags as the rows and used the <td> tags as the individual cells of data. At times the tables that contained the information that we needed to pull was within comments on the HTML. In this case we had to simulate a user using the webpage and read through the comments of the webpage. Once this was performed we were then able to pull all of the data that we needed. With the help of Pandas we saved all of the pulled data in excel files as Dataframes. This allowed for easy manipulation and merging of each table.

# 4. Design

# 5. Results

## Evaluating the Evaluators: Analyzing Player Performance Metrics

As discussed in the methods section, a crucial decision in evaluating player value is how ‘performance’ is quantified. The below table lists the top 20 players ranked using our six selected metrics.

Table 2: Top 20 Table by Metric

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Win Shares** | **Player Efficiency Rating** | **Value Over Replacement Player** | **Basic Percentile** | **Advanced Percentile** | **Fantasy Points** |
| Lebron James  Karl Malone  Michael Jordan  David Robinson  Chris Paul  John Stockton  Kevin Durant  Tim Duncan  Magic Johnson  Charles Barkley  Reggie Miller  James Harden  Dirk Nowitzki  Shaquille O’Neal  Stephen Curry  Clyde Drexler  Kevin Garnett  Dwight Howard  Jeff Hornacek  Kobe Bryant | David Robinson  LeBron James  Shaquille O’Neal  Karl Malone  Charles Barkley  Tim Duncan  Chris Paul  Kevin Durant  Yao Ming  Hakeem Olajuwon  Michael Jordan  Dwyane Wade  John Stockton  Anthony Davis  Kobe Bryant  Russell Westbrook  Kevin Garnett  Dirk Nowitzki  Clyde Drexler  Amare Stoudemire | LeBron James  Karl Malone  Michael Jordan  David Robinson  Chris Paul  Charles Barkley  Clyde Drexler  Russell Westbrook  Tim Duncan  Steph Curry  James Harden  Kevin Garnett  Scottie Pippen  Larry Bird  Kevin Durant  Jason Kidd  Hakeem Olajuwon  Shaquille O’Neal  John Stockton  Dwyane Wade | Chris Webber  Larry Bird  Hakeem Olajuwon  Karl Malone  Clyde Drexler  David Robinson  LeBron James  Kevin Garnett  Charles Barkley  Scottie Pippen  Tim Duncan  Dwyane Wade  Kevin Durant  Kobe Bryant  Chris Bosh  Shawn Marion  Jason Kidd  Larry Nance  Lamar Odom  Vlade Divac | David Robinson  Hakeem Olajuwon  Karl Malone  Kevin Garnett  Charles Barkley  Vlade Divac  Tim Duncan  LeBron James  Larry Bird  Scottie Pippen  Chris Webber  Shawn Kemp  Shaquille O’Neal  Clyde Drexler  Christian Laettner  Paul Millsap  DeMarcus Cousins  Larry Nance  Dwyane Wade  Marcus Camby | LeBron James  Karl Malone  Shaquille O’Neal  Tim Duncan  David Robinson  Russell Westbrook  Kevin Durant  Kevin Garnett  Allen Iverson  Hakeem Olajuwon  Chris Paul  Dwight Howard  Kobe Bryant  Gary Payton  Jason Kidd  John Stockton  James Harden  Dirk Nowitzki  Steph Curry  Patrick Ewing |

PER and VORP similar? Only WS and VORP consider anything defensive. Advanced Percentile has some pretty different players in the top 20 (Divac, Laettner, Millsap, Cousins, Nance, Camby).

Discussion about the table. Unique players (ones that only appear once). Talk about outliers, add them to the sheet.

## Average NBA Draft Pick Value

## Predicting NBA Performance for NCAA DI players

# 6. Discussion and Future Work

# References