

Drafting Errors and Decision Making Theory in the NBA Draft

Daniel Sailofsky, LL.B

Submitted in partial fulfillment of the requirements for the degree of
Master of Arts in Applied Health Sciences
(Sport Management)

Faculty of Applied Health Sciences

Brock University

St. Catharines, Ontario

©April 6, 2018

Abstract

Even with the recent influx of available data with regards to draft-eligible players and NBA teams investing more resources into scouting than ever before, NBA decision makers still struggle to consistently evaluate talent and select productive players (Berri et al., 2010) in the draft. In this paper, I examine the NCAA statistics and pre-draft player factors that predict both draft position and NBA performance for all NCAA players drafted to the NBA between 2006-2013. Following this analysis, I determine what errors NBA teams are making and how these errors relate to general decision making theory.

To compare the predictors of draft position and NBA performance, linear regression models are specified for both draft position and NBA performance. The NBA performance model sample necessarily excludes players whose production cannot be assessed due to not playing a minimum (≥ 500) amount of NBA minutes, and therefore a Heckman (1971) sample selection correction is applied to the performance model to correct for this non-randomly selected sample. Both models are specified for the entire dataset as well as for subsets for position (Big, Wings, Point Guards) and conference size (Big Conference, Small Conference).

The findings of this paper demonstrate that NBA decision makers continue to base their draft selections on factors that do not actually predict future NBA success, such as scoring, size, and college conference. Many of the decisions made by NBA decision makers relate to Heath and Tversky's (1991) competency hypothesis, as front offices forego the use of reliable distributive data and select players according to their perceived knowledge. NBA decision makers also display risk averse behaviour (Kahneman and Tversky, 1973) and an insistence on sticking with the status quo (Samuelson and

Drafting Errors and Decision Making Theory in the NBA Draft

Zeckhauser, 1988) in their decisions. More specifically, this study also points to ball control and offensive efficiency as predictors of individual player success. These findings can not only affect NBA decision makers in the factors that they emphasize in player evaluations, but can also be used to change the way that sport executives think about general decision making and their own innate decision making biases.

Keywords: NBA, draft, decision making, regression, Heckman

Table of Contents

Abstract	
1. Introduction	1
1.1. Purpose	5
2. Literature Review	5
2.1. Amateur Player Drafts in North American Sport.....	6
2.2. Statistical Models for Players’ Effect on Team Success.....	20
2.3. Statistical Models in NCAA Basketball	24
3. Background	32
3.1. History of the NBA Draft.....	33
3.2. NCAA Division I Basketball.....	34
4. Methodology	38
4.1. Theoretical Approach.....	38
4.2. Empirical Approach.....	46
5. Data Description	49
5.1. Data.....	49
5.2. Variables.....	51
6. Results	58
6.1. Summary statistics	58
6.2. Full Model Estimates	60
6.3. Supplementary position based model.....	63
6.4. Supplementary Conference Based Models	69
7. Discussion	71
7.1. Drafting Errors and Decision Making Biases	71
7.2. Specific Statistics in the General Models.....	77
7.3. Supplementary Model Specific Biases	87
8. Conclusion	91
9. Tables	95
10. Code	109
10.1. Code for creating Master spreadsheet from individual player CSVs	109
10.2. Code for Updating the Master Spread Sheet.....	112
10.3. Code for Summary Stats	117
10.4. Code for non-Heckman analyses (Python)	118
10.5. Code for Heckman corrected analyses (R).....	125
10. References	128

1. Introduction

Prediction and decision making is not easy, and nor is it simple or always rational. In the 1970's, Kahneman and Tversky's (1971; 1978; 1979) groundbreaking decision making research changed the way many thought about decision making and prediction, as they highlighted the implicit biases that explain the reason human beings make certain decisions. Whether it is trying to pick the choice with the highest expected utility (Kahneman and Tversky, 1979) or attempting to predict the results of sporting events (Heath and Tversky, 1991), people often suffer from systematic errors and biases in their decision making, which can effect their ability to make rational choices.

Prediction and decision making come into play in a variety of different employment contexts, including professional sport. The amateur player draft represents an especially fitting research context to examine decision making, as front office members use a variety of different criteria to evaluate and select amateur players, with the goal of improving their team. This is akin to employers evaluating and hiring employees, as it requires decision makers to predict the future success of their hires based on past performance and other personal factors that they deem important.

This study focuses on decision making in the National Basketball Association (NBA) by examining the selections of NBA decision makers in the league's amateur player draft. To do this, NCAA Division I basketball performance statistics and pre-draft measurable characteristics will be assessed to determine which factors predict where a player will be selected in the draft as well as their future production at the NBA level.

The NBA draft is an interesting testing ground for decision making, even more so because of how important the draft is for NBA teams themselves. For many teams, the draft represents the best chance they have at improving. Unlike free agency, where players take into account a multitude of factors including the players and coaches already on a team, the history and prestige of that team, and even the city life, state taxes and the weather of their new prospective homes before making their decision, it is the teams that are in control at the draft. Given the value that individual players can have in a basketball game - where players can play more than 75% of the game and effect their teams' chances of winning in a more substantial way than in sports such as football, baseball or hockey – acquiring talent at the draft is of the utmost importance for NBA teams (Sanderson and Siegfried, 2003). The NBA's collective bargaining agreement also limits the salaries of rookies for up to four years (Sadler and Sanders, 2016), which has made impactful rookies and young players on rookie contracts even more valuable assets, as they can deliver production at a much lower pay grade than their veteran counterparts (Berri, 2012).

Due to NBA rules and the traditional structure of amateur basketball in North America, most drafted players come from NCAA Division I programs. In the 2016-2017 NBA season, 80.2% of players on opening day rosters played at NCAA Division I colleges (Sukup 2017). While there are players who play in the professional ranks internationally before entering the NBA draft, these players are not included in the analysis. This is due to the fact that it would be too complex to compare the statistics that these players record at various international levels with those recorded by players at the NCAA Division I level. Comparing players from just one international league with

college players would be difficult enough, and to add to this, international league players come from a variety of different leagues that can vary considerably in their quality of play and style. This is discussed in further detail in Chapter 2.

Though teams are aware of the importance of the draft and invest heavily in scouting and player evaluation resources, drafting has proven to be extremely difficult (Berri et al., 2011). NBA history is littered with high draft picks and former NCAA stars who have failed to produce at the next level, as well as lower draft picks and overlooked players who have thrived (Groothuis et al., 2009). There is little correlation between the pick a player is selected with and their future success in the NBA, as illustrated below. This has held true in other professional sports leagues as well, as researchers have found that these drafting errors can be explained by several decision making biases. These include overconfidence, non regressive predictions, and biases in favour of players from higher profile colleges in the NFL, as well as biases towards high school players in MLB and biases against European, French-Canadian, and smaller players (in terms of both height and weight) in the NHL.

The motivation for this study is thus to examine NBA decision making by examining which pre-NBA factors act as predictors for draft position and which act as predictors for NBA success. This determines whether NBA decision makers are drafting optimally, and if they are not, what decision making biases are effecting these draft-day decisions. While some research has been done to assess the relationship between draft position and performance, this past research has grouped players using outdated position types and has not had access to the multitude of advanced basketball statistics now available that more accurately assess NCAA and NBA performance. Instead of simply

using traditional box score statistics or rate based versions of these statistics, we replace these with percentage based statistics that more accurately assess a player's personal contributions to winning given the opportunities they are presented with. Percentage based statistics how often a player contributes in a specific statistical category based on the opportunities they have. Using these statistics (when possible) allows for an assessment of a player's statistical output that minimizes the impact of factors out of a player's control, such as playing time and offensive and defensive pace. For example, rebound percentage measures the percentage of available missed shots a player rebounds while they are on the court, rather than simply calculating the total rebounds a player accumulates on a per game or per minute basis. Past work has also failed to account for the sample selection issue (discussed in detail below) that can arise in research of this kind.

Extra information regarding draft eligible NCAA players is valuable for NBA front offices when making decisions, especially with regards to a talent pool that has proven to be extremely volatile and difficult to assess (Berri et al., 2011). Scouting the way it is currently done is still far from an exact science and can be improved. An objective of this research is thus to use statistical modelling to provide insight and information regarding college players' future success in the NBA. Empirical sample selection and regression models will be constructed to isolate the NCAA statistics and pre-draft factors that predict NBA success and draft pick selection for the 372 players drafted from NCAA Division 1 basketball between 2006-2013.

The results of this analysis can help teams and their decision makers add an extra layer to their scouting of potential draft picks. Moreover, by examining position specific

and conference specific predictors, we can uncover front office errors and biases related to specific positions and player types, which can be a great benefit to many players whose skill sets and value have traditionally been overlooked. This can also highlight decision making patterns made by front offices with regards to certain types of players, which in turn could have a positive effect on these players' future draft stock. These biases could also be generalizable to other sports leagues or decision making contexts.

1.1. Purpose

The objectives of this research are:

1. To determine the NCAA Division 1 basketball performance statistics, team factors, physical characteristics, specific conference participation, and playing position that predict draft position.
2. To determine the likelihood of NCAA Division 1 basketball players drafted to the NBA playing in the NBA.
3. To determine the NCAA Division 1 basketball performance statistics, team factors, physical characteristics, specific conference participation, and playing position that predict NBA performance among drafted players.
4. To examine the potential disjuncture between the findings of objectives 1 and 3, in order to determine what decision making biases exist in NBA front offices when drafting NCAA players.

2. Literature Review

In this chapter, I examine the literature with regards to amateur player drafts in North American professional sports. This research includes work that examines the

difficulty and importance of drafting productive players in professional sports, as well as the NBA draft and the incentives for teams to lose games to better position themselves. I also study the Sunk Cost Effect observed amongst NBA decision makers. Next, I turn to the relevant literature surrounding player performance models and the individual statistics and skills that lead to team success, which includes research done with regards to the NBA as well as other professional and amateur basketball leagues. Finally, I examine the literature surrounding predictive statistical modelling in the NCAA. While most research done in this area revolves around constructing win probability models for the season ending NCAA Tournament, this section will include academic literature modelling how players' NCAA statistics and pre-draft factors can be used to predict where they will be drafted as well as certain aspects of their NBA performance and careers.

2.1. Amateur Player Drafts in North American Sport

The NBA, like their counterparts in the three other major North American professional sports leagues, uses a reverse order draft to disperse incoming amateur talent. In a strict reverse order draft, the draft position of each team is based on the reverse order of the previous year's standings. While there are slight variations in each the four professional leagues in how they allocate their top picks, the NBA, National Football League (NFL), National Hockey League (NHL), and Major League Baseball (MLB) all conduct a version of this type of draft at season's end (Berri et al, 2011).

2.1.1. The difficulty of drafting in the NBA and professional sports. The NBA draft has proven to be a difficult process. After the first overall selection, the chances of drafting a "superstar" - defined as a player with at least one season where their performance was two standard deviations above the mean – drops drastically (Groothuis

et al. 2009). Several other studies (Coates and Oguntimein, 2010; Berri et al., 2011; Harris and Berri, 2015) that have assessed the statistics and factors that affect future NBA performance and draft slot also illustrate the inability of decision makers to draft optimally. For the purposes of this manuscript study at large, “drafting optimally” refers to drafting players who perform at the highest level in terms of helping their teams win at the professional level, regardless of salary considerations.

Koz et al. (2011) examined the difficulty of drafting in all four of the major North American professional sports leagues. Using games played as the common performance proxy variable across all four sports, they found that only 17% of the total variance in players’ games played was due to where they were drafted. For the NBA, NFL and NHL, draft round was correlated with games played, though the authors mentioned that this could be due to the Sunk Cost Effect studied by Staw and Hoang (1995) and Keefer (2015) with regards to NBA and NFL draft picks. This sunk cost effect is discussed in detail below, but is generally based on the idea that teams are more likely to grant their higher draft picks more playing time and opportunities even when their performance does not warrant it. Keefer (2015) conducted a similar study with regards to NFL draft picks and found the same effect. It is also important to note that Koz et al. (2011) used draft round as the independent variable rather than the specific pick number. While this may be a useful measurement for other sports leagues, the NBA draft’s current two-round structure creates a wider gap between 1st and 2nd round picks and even between players selected earlier in a round versus later. Koz et al. (2011) found nearly no relationship between draft round and games played for MLB draftees.

2.1.2. The NFL draft. Massey and Thaler (2013) determined that NFL decision makers vastly overestimate their own abilities to evaluate talent and make draft choices. They found that the NFL draft is still vital for the success of franchises, but that selecting players is extremely difficult. NFL front offices did not consistently select productive players in higher draft slots, to the point that lower round picks (2nd, 3rd, etc.) were more valuable than 1st round selections due to the higher salary structure designated for these higher round picks. Teams nevertheless consistently inflate the value of 1st and 2nd round selections, which resulted in situations where teams that traded down by trading their first round picks for two lower round picks were actually acquiring *two* superior picks to the one they traded away. In the case of the NFL draft, Massey and Thaler (2013) found that evaluating talent is so difficult and there is so much volatility in selections that simply having as many picks as possible (rather than fewer higher selections) is the best option for acquiring talent. Hurley et al. (2012) also looked at the value of draft picks and came to similar conclusions regarding the value of certain selections.

Berri and Simmons (2011) found little correlation between where NCAA quarterbacks were drafted and their future NFL success. Moreover, the statistics and qualities that predicted where a quarterback was drafted were not the same as those that actually predicted performance, demonstrating a disjuncture in the use of available information by NFL decision makers. The only strong relationship between performance and draft slot for NFL quarterbacks was between draft slot and playing time, again pointing to teams' insistence on trusting their own processes and playing the players they drafted. Berri and Simmons (2011) finished their study by highlighting the difficulty of predicting professional performance based on statistics and factors from the amateur

collegiate level. Wolfson et al. (2011) also noted the difficulty of drafting and predicting future performance in their analysis forecasting NCAA quarterback performance in the NFL.

Boulier et al. (2010) found that, while they are far from perfect, NFL front offices are successful in evaluating the future performance of collegiate quarterbacks and wide receivers. This study was conducted on all players drafted between 1974-2005. It is important to note that the proxy statistics used to measure performance for quarterback play included two longevity and playing time based measurements. Boulier et al. (2010) base performance heavily on of the number of years a quarterback played as well as how many total passes they threw. These are statistics that are based on how much a player is used, and can again be prone to NFL decision makers exhibiting Sunk Cost biases (Keefer, 2015, Staw and Hoang, 1995). For wide receivers, the total number of games and seasons played by a player are also used as proxies for performance. The only dependent variable statistics used in their analysis that are not based solely on how much a player plays are Quarterback Rating for quarterbacks and receiving yards for wide receivers. While quarterback rating does isolate a player's performance and does not advantage those who are provided with more opportunities, receiving yards is a raw total statistic that again rewards players who are given more playing time.

Hendricks et al. (2003) looked at NFL players in an employment selection context in order to assess their production and how these "assets" are picked based on the pre-NFL information available. They found that NFL teams made systematic errors with regards to players from more visible, higher profile colleges. The abilities of players from historically higher performing colleges from the traditional power conferences such as the

SEC, Big 10, ACC and Pac-10 were consistently overrated by NFL decision makers. Hendricks et al. (2003) also found a league wide tendency towards risk aversion, even though this behaviour is detrimental to most teams. Similarly, Kitchens (2015) determined that individual players from successful NCAA teams were drafted higher in the NFL draft, even though playing for one of these schools had no significant impact on a player's NFL performance.

Finally, the NFL draft combine - where players perform speed, agility and weightlifting drills in front of NFL decision makers - provides even more pre-draft information for NFL decision makers. However, this added information has not typically resulted in more optimized decision-making (Kuzmits et al., 2008; Robbins, 2010; Berri and Simmons, 2011).

2.1.3. The MLB draft. As the first of the four North American sports to embrace data analytics, many researchers have written about the MLB draft, the labour market, and the productivity and salary allocation of its players. Only the most relevant draft research is examined here. The MLB draft is the largest of any of the four North American professional sports, with 50 rounds and 1500 players selected every year. With no hard salary cap in the sport, the draft takes on added importance as it presents small-market, lower revenue teams with the chance to draft talented young players and keep them on lower, cost controlled salaries for the beginning of their careers (Streib et al., 2012). However, like in the NBA and NFL drafts described above, teams still do not always draft optimally.

Spurr (2000) found that there was little difference between teams' abilities to find MLB level talent in the draft. He also found that American college players were more

likely than American high school players to play and be successful in the MLB, though MLB decision makers “failed to recognize the superior prospects of the college and elite college player for a substantial period of time—for at least 3 years and perhaps much longer” (Spurr, 2000, p. 79). However, Caporale and Collier (2013) found that for the very top of the draft (the first round), there is no advantage for college players. They determined that for five years worth of first round draft picks, there was no significant difference in future MLB production between players drafted from high school or college.

Burger and Walters (2009) also examined teams’ assessment of high school and college players. They used data on draftee bonuses and future performance to determine that teams overvalued high school players relative to college players. Teams also overvalued pitchers relative to position players. Burger and Walters (2009) hypothesized that this overvaluation could be due to teams’ overestimation of how often high school players become stars relative to college players. In reality, this difference is negligible and is outweighed by the number of high school players who do not end up playing in the MLB at all and also by how much sooner college players contribute at the MLB level due to their advanced age and maturity at the time they are drafted. The magnitude of these inefficiencies point to the existence of “deep-seated biases in decision making” (p. 498) among MLB front offices.

Several legal scholars (Bailey and Shepherd, 2011; Hauptman, 2010; Poydenis 2010) wrote about the racist effects of the MLB draft and the need for a global draft. Since the 1960’s, the number of Latin-American players populating MLB rosters has grown tremendously. All MLB teams currently have baseball academies in the

Dominican Republic, while most have in Venezuela and other Latin-American countries as well (Bailey and Shepherd, 2011). The amateur draft, however, is still limited to North American players. This results in a system where there is no pay scale for young Latin American players, and many of these players can be signed for lower initial bonuses than they would have received in the draft (Bailey and Shepherd, 2011).

2.1.4. The NHL draft. NHL front offices are prone to decision-making errors in their draft as well. Hurley et al. (2012) analyzed data regarding 4496 drafted players from 1980 to 1997, and found that there is more variation in the eventual successes of NHL draft picks than for NFL draft picks. This could be due to the fact that NHL players are normally drafted at 18 years old, the youngest of any of the four North American professional sports and much younger than NFL players, who are normally drafted at 22-23. Players drafted at an earlier age are less physically mature, and there is also less available data to analyze their performance and more uncertainty regarding how the athlete will develop.

Many researchers have examined entry draft discrimination for different nationalities and player types, though only the most recent and relevant analyses are listed here. Studying players from the 1983-1984, 1989-1990, 1990-1991, and 1993-1994 seasons, Lavoie (2003) found front offices hold biases against French Canadian players, especially American teams. The only teams that did not show this bias were from Montreal and Quebec City, two cities with obvious French Canadian ties. Christie and Lavoie (2015) also analyzed player performance and draft position from the rosters of teams in the 1993-1994 and 2009-2010 seasons to determine whether draft biases for other player nationalities existed and still exist. Data from the 1993-1994 season showed

that biases against French Canadian players still existed at that time. However, these biases disappeared when assessing players in the 2009-2010 season, demonstrating the evolution and improvement of NHL team decision making through the years. Christie and Lavoie (2015) also found clear biases against the drafting of European players when looking at the 1993-1994 rosters. This coincides with research done by Dawson and Magee (2012), who found that European players drafted between 1988-1992 outperformed their draft rank. However, examining data from more recent years, Christie and Lavoie (2015) found that this European player bias has diminished. The only nationality which NHL decision makers still exhibit a bias towards with regards to Russian players. This is due in large part to the existence and importance of the Kontinental Hockey League (KHL) to Russian players, as NHL front offices are hesitant to draft players who are either already locked into KHL contracts or have a higher chance of staying in Russia and not making the jump to the NHL.

Voyer and Wright (1998) conducted a regression analysis using Junior hockey level performance statistics, NHL statistics, and draft data regarding all 740 players who played at least one game during the 1993-1994 season. The goal of this study was to determine which Junior hockey factors predict future NHL performance, as well as to assess whether NHL decision makers are effective in their selection of Junior players. Voyer and Wright (1998) used NHL regular season points scored per game as well as NHL playoff points scored per game as the dependent variables to represent NHL performance in their regressions. They used Junior hockey performance statistics, entry draft pick slot, and height and weight when drafted as the independent predictor variables. They found that regular season points scored in Junior was the best predictor of

NHL success, followed by awards won and goals scored in Junior. In terms of position specific predictors, only regular season points scored in Junior acted as a predictor of NHL regular season and playoff performance for defenseman, while a low number of penalty minutes acted as a predictor of NHL regular season performance. Right-handed left wingers were also found to be more successful NHL players in both the regular season and the playoffs. There were several other minor differences between the different forward positions with respect to their predictors of NHL performance as well.

Voyer and Wright (1998) found that size (both height and weight) was not an important predictor variable of NHL performance, though it has traditionally been treated as an important factor in the evaluation of players. This overestimation of the importance of size is perhaps one of the reasons for the difficulties NHL general managers have in drafting effective players. This inability is reflected in Voyer and Wright (1998) as well, as they found that “entry draft rank tended not to be a useful predictor of NHL performance” (461).

Dawson and Magee (2012) examined NHL draft data and the NHL statistics of players drafted between 1969-1995 to assess the relative value of each draft pick as well as evaluate the decision making abilities of NHL franchises. Unlike Voyer and Wright (1998), Dawson and Magee (2012) measured NHL performance according to games played in the NHL, using this as the dependent variable in their regressions. Using this criteria, they found that Buffalo, Quebec City/Colorado, Winnipeg/Phoenix and the New York Rangers have drafted better than other teams. These teams more consistently drafted players who ended up playing more games in the NHL. They also found more variation for the performance of goaltenders selected early in the draft than for forwards

and defenseman, which could explain why relatively fewer goaltenders are selected in the first round.

Tarter et al. (2009) utilized a different approach to examine the factors that predict future NHL performance than the above mentioned researchers, as they focused on 345 players' strength and conditioning test scores at the NHL draft combine. The authors measured these players' scores and developed an index score for their total combine performance. They found that players measuring in the 90th percentile or higher in this index score were 72% (forwards) and 65% (defenseman) more likely to play in the NHL within four years of the draft than those below the 90th percentile.

2.1.5. The NBA draft and the incentive to lose. The importance of the NBA draft is evidenced by how teams have reacted to and continue to react to its structure and the way the draft order is set. Before 1985, the NBA operated under a strict version of the reverse order draft, where pick order was based directly on the reverse order of the previous year's standings. This encouraged teams to lose games at the end of the season once they realized that their chances of making the playoffs or winning a championship were slim (Price et al. 2010). Price et al. (2010) determined that this effect was even more pronounced as teams' chances of making the playoffs continued to decrease. Taylor and Trogon (2002) noted that in these instances, NBA teams were simply responding to tournament incentives, as they were more likely to lose games when they were incentivized to do so. Soebbing and Mason (2009) found even more support for this incentive to lose phenomenon. When a generational talent was set to declare for that year's draft, teams were even more likely to lose games to increase their chances of drafting that player (Soebbing and Mason, 2009).

This trend forced the NBA to alter their draft lottery system in 1985. In the 1983-1984 season, the year before this change, Taylor and Trogdon (2002) observed that teams eliminated from playoff contention were 2.5 times more likely than noneliminated teams to lose games, even when controlling for team quality. A year later, the draft lottery system was altered to give all nonplayoff teams an equal chance at the top selections and the draft order was decided as a pure lottery amongst these teams. When this change occurred in the 1984-1985 season and the incentive to drop to the bottom of the standings disappeared, Taylor and Trogdon (2002) found no significant difference between eliminated and noneliminated teams. After the 1989-1990 season, the league reformatted the draft lottery again to its current system, which is a compromise between the strict reverse order draft and a true draft lottery. In this system, all non playoff teams are given a weighted chance - relative to the reverse order of the previous season's standings - at one of the top three picks, with the rest of the selections following in the reverse order of the standings. This reintroduced the incentive to lose games in order to increase one's chances at a higher selection, and with this, eliminated teams again began to lose more often. In the 1989-1990 season, the first with the current draft lottery system, eliminated teams were once again almost twice as likely to lose games relative to noneliminated teams (Taylor and Trogdon, 2002).

2.1.6. The importance of star players in the NBA. Research has also shown that the importance of drafting star players is especially pronounced in the NBA, which could result in decision making biases that favour high risk, high upside players in the NBA draft. With basketball teams only playing five players on the court at a time and individual players able to play 75% or more of each game, the value and impact star

players can have is larger than in other team sports (Sanderson and Siegfried, 2003). This further incentivizes teams to lose games and increase their chances of drafting a great player at a higher draft position. Research regarding the effect individual players have on their teams also shows the importance of acquiring productive players. Berri (2010) determined that 80% of the Wins Produced - a metric developed to assess players' individual contributions to team wins - on an NBA team comes from 22.6% of its players. In other words, on an average NBA team of 16 players (though NBA rosters are now capped at 15), 80% of that team's wins will be produced by 3-4 players (Berri, 2010).

The small labour markets of professional sports leagues further augments the importance of acquiring higher draft picks and using these picks effectively. Schmidt and Berri (2003) looked at the effect that the MLB labour market has on competitive balance. They found that the smaller labour market of the MLB negates the purported effect that the draft and free agency have on competitive balance. In professional sports, there are a limited number of players who can change a team's fortunes, and this effect is further exacerbated in the sport of basketball. Berri et al. (2005) commented on the proliferation of this phenomenon in the NBA, where there is a smaller labour market compared to other professional sports leagues and thus significantly less competitive balance. This smaller labour market is due in large part to the importance of height in the sport. 97.9% of young adult males in the United States are six feet three inches or shorter, but less than 20% of NBA players in the 2003-2004 season fit this description. In the same vein, 30% of NBA players are six feet ten inches or taller, which is a very difficult characteristic to find in the general population. With this dearth of available tall people, there is also more

variability in the few tall peoples' talent levels. This results in a much smaller pool of potential players and especially potential franchise changing players from which to pick, further highlighting the importance of acquiring higher draft picks (Berri et al., 2005).

There is a financial incentive for NBA franchises to acquire star players as well. Berri et al. (2004) looked at the impact that a player's All Star Game appearances and All Star votes have on team revenues, and found a statistically significant increase in gate revenue and attracting fans to games. Though team wins had an even stronger positive relationship with team revenues, productive All Star level players help their teams win and therefore help in this regard as well (Berri et al., 2004). Hausman and Leonard (1997) found that the strongest direct effect that All Star players have on gate revenue is actually for opposing teams, as more fans will attend games where they can see a visiting All Star player that they do not normally get to watch.

2.1.7. The Value of Rookie Contracts. With changes to the NBA's collective bargaining agreements in 2005 and 2011 limiting rookie pay scales (Berri, 2012), the value of drafted players on entry level rookie contracts has risen considerably. Regardless of how much money a first year player could demand on a theoretical open market, all rookie contracts are limited by the draft position of the player, with *maximum* first year salaries ranging from 1.16 million for the 30th pick to 5.85 million for the 1st pick in the draft (NBA Rookie Scale 2017-2018). In the 2010-2011 season, for example, Kevin Love produced wins at a level worth 32.23 million dollars while playing on his 3.63 million dollar, third year rookie salary (Berri, 2012). While most players on their rookie contracts will not be able to help produce wins at quite this level, their value is still higher due to this restricted salary scale. Moreover, for their second contracts, teams can offer more

money and longer term deals to players that they have drafted and who have stayed with the organization for the length of their rookie contracts, making the selection of productive players in the draft even more important (Sadler and Sanders, 2016).

2.1.8. The NBA's Sunk Cost Effect. The Sunk Cost Effect studied in economics has been shown to apply to high NBA draft picks as well (Staw and Hoang, 1995). When a player is selected with a high pick in the draft, their team is more likely to continue to give them more playing time and chances to prove themselves, even when they show no signs of being a productive player. Teams seemingly want to validate their own selections, to the point that when they make a poor choice, the negative consequences of that choice continue to plague them as they keep using the player to the detriment of their own team (Staw and Hoang, 1995). Keefer (2015) found the same effect among National Football League (NFL) front offices and their draft picks, as teams were more likely to give more playing time and more lucrative contracts to first round draft picks even when their production did not merit it.

Camerer and Weber (1999) re-examined Staw and Hoang's (1995) results with a new dataset while controlling for other possible explanations for a player's playing time. Their research showed a slightly smaller escalation effect than that observed by Staw and Hoang (1995), but their results were nevertheless significant in demonstrating the same Sunk Cost Effect for players drafted higher. Groothuis and Hill's (2004) analysis of NBA career duration and length also showed that, controlling for the effects of player performance and physical characteristics, where a player was drafted had an effect on the length of their careers. Players selected early in the draft seemingly got the benefit of the

doubt throughout their careers, to the point that it affected how long they were able to stay in the league (Groothuis and Hill, 2004).

This reluctance to admit errors in drafting is linked with Massey and Thaler's (2005) research on the NFL draft as well. Their analysis showed that NFL front offices overestimate their own ability to evaluate talent and make optimal draft choices. Professional sports general managers and front offices often believe that their ability to evaluate talent and make these decisions is what separates them from their peers, and this overconfidence can result in drafting mistakes that are compounded through an insistence that the player they selected will eventually produce as they initially expected (Massey and Thaler, 2005). This is yet another example of the systematic decision making errors made by NBA general managers and front office members.

2.2. Statistical Models for Players' Effect on Team Success

Next, I provide an overview of the predictive models used to determine the way different individual and team basketball statistics effect team success. Researchers have produced statistical models to predict basketball performance, though not often in terms of identifying NCAA basketball statistical predictors for NBA success.

2.2.1. NBA Performance Models. Researchers have attempted to identify the player and position types that contribute most to winning in NBA basketball. Examining NBA regular season data from 1994-1998, Berri (1999) determined that a "player's production of wins is primarily a function of his ability to acquire and maintain possession of the ball, and his ability to convert his field goal attempts consistently" (p. 417). In other words, a player's contribution to winning derives from his ability to acquire rebounds while avoiding turnovers and shooting efficiently. In later research

examining a different sample of NBA seasons, Berri and Schmidt (2006) again found that the individual player statistical factors that most affect winning were shooting efficiency, rebounding, and steals.

Lee and Berri (2008) analyzed NBA players' contributions to team success to determine which players and which position groups had the strongest effect. They examined player data from 1999-2000 to 2001-2002 and grouped players by position. To assess individual players' contributions to winning, Lee and Berri (2008) based their model on the premise that wins in the NBA "are determined by how efficiently one scores per possession employed, relative to one's opponent's ability to use possessions efficiently" (p. 53). They therefore measured player production according to a variety of shooting efficiency (i.e. field goal percentage), ball control (i.e. turnovers) and ball retrieval (i.e. rebounds) statistics in order to determine which players and position types had the strongest effect on winning. In incorporating the position groups into their analysis, Lee and Berri (2008) found that power forwards and centres were most valuable.

Page et al. (2007) examined players' box score statistics (adjusted for 100 possessions) and their effect on teams' ability to win individual games. They used the point differential of each game in the 1996-1997 NBA season as the dependent variables in their analyses. They then collected box score statistics for each position on both teams, and used the difference between the position pairings as independent explanatory variables. For example, if the home teams' power forward scored 16 points per 100 possessions while the away teams' scored 13, then for this particular game, the home team power forward would have a +3 score in the points per 100 possessions variable.

This calculation was conducted for each box-score statistic and these statistical differences were used as explanatory variables. The results of these analyses showed that for all five positions (point guard, shooting guard, small forward, power forward, centre), the positive point differential for a team increased if that team's players had more offensive rebounds and assists, shot a higher percentage on field goals, and had less turnovers than their positional counterpart. Defensive rebounding was only significant for the shooting guard position, while offensive rebounding had an especially pronounced effect for point guards. Simply scoring more points than your positional counterpart was not significant on the team point differential, while field goals made was only significant for centres. With regard to all of these results however, it is important to note that this analysis only takes into account the players who started the game for each position, as Page et al. (2007) compiled the contributions of all other players into a separate positional grouping called "bench". The results for specific position groups were thus based only off of the box score statistics of the starting players.

Deshpande and Jensen (2016) examined the effect players had on winning in terms of their plus/minus scores when they were on the court. They did this by measuring score changes over each player's time spent on the court, while placing added or reduced value on a player's contribution depending on the score and time left in the game. In short, minutes played when the score was tied or very close late in a game were weighted more heavily than those played in a blowout. By looking at a player's box score totals and how much their team's win probability increased or decreased between the time a player began their shift and when it ended, Deshpande and Jensen were able to determine an Impact Score for each player.

2.2.2. Performance models from other basketball leagues. Researchers also looked at game related individual statistics and their effect on winning for other professional and amateur basketball leagues. Sampaio et al. (2006) used a position specific analysis of player statistics in the NBA and in other professional leagues to try to understand the importance of different individual skills to winning at the team level. They determined that there is a large difference between the statistics that affect winning for each position type (guards, forwards, centres). There is also a notable difference in the importance of certain statistics based on the quality or style of play in a particular league (Sampaio et al., 2006).

Like Sampaio et al. (2006), Choi et al. (2015) noted the varied importance of certain statistics for different position types in the Korean Basketball League. They also further highlighted the effect that cultural and environmental factors have on the style and quality of play in a league, and the way this subsequently effects the importance of certain skills and statistics for players in that league. Ibanez et al. (2008) conducted a similar study with all players in the Spanish Basketball League (without controlling for position), looking at which individual statistics best predicted season-long team success. They determined that assists, steals and blocks were the most accurate season long individual predictors of team success at this level. On a team statistic level, Ibanez et al. (2008) found that offensive shooting efficiency had a positive relationship with winning, while a faster or slower pace of play did not have any significant effect.

Koh et al. (2011) examined Youth Olympic basketball players to determine what differentiates players on successful teams compared to those on unsuccessful teams. They found that being taller, shooting a higher percentage on field goals, and playing more

aggressively by both drawing opponents' fouls and committing more team fouls were determinants of team success.

2.3. Statistical Models in NCAA Basketball

Most work done analyzing individual player statistics for NCAA basketball has been in the form of NCAA Division 1 Basketball Tournament models that attempt to predict the winners of tournament games. With the popularity of the NCAA Tournament, almost all individual player statistical analyses have been used as a tool to predict teams winning at the college level rather than how these statistics will affect a college player's ability to positively affect his future NBA team.

2.3.1. NCAA player prediction models for players drafted to the NBA. Only a few studies specifically analyzed college statistics and their effect on the NBA draft and NBA careers, though the data now exists at the college level for more statistical analysis (Berri, 2005). Kahn and Scherer examined the statistics that had the strongest correlation with high draft positions in order to assess whether general managers and front office biases exist when drafting players (Kahn and Scherer, 1988). This study did not look at the relationship between these pre-draft variables and future performance, but rather whether these factors eventually effected compensation and whether compensation varied by race.

Abrams et al. (2008) examined the relationship between NCAA statistics, pre-draft variables and NBA career longevity, though not necessarily production. While career longevity or playing time can sometimes act as a proxy for success (Massey and Thaler, 2011; Wolfson et al., 2011; Rodenberg and Kim, 2012), this study did not directly examine a player's production at the NBA level in terms of their effect on winning.

Abrams et al. grouped players by the three traditional position types (guards, forwards and centres) and used their college statistics and pre-draft factors as independent variables. The number of total NBA seasons played was used as the dependent variable. They determined that assists, steals, turnovers and points scored had a significant relationship with career longevity for guards, while field goal percentage, free throw percentage and assists showed a similar significant relationship for forwards. Interestingly, no significant relationships were found between the pre-draft variables and NBA career longevity for centres (Abrams et al., 2008).

Coates and Oguntimein (2010) analyzed the NCAA statistics that translated to the NBA level. They did not group players differently according to their positions, but did control for the NCAA conference the player played in. They grouped drafted players in two groups according to the conference they played in, with players from one of the top eight performing conferences in a Big Conference group and all other players in a Small Conference group. Interestingly, Coates and Oguntimein (2010) noted that while Small Conference players were drafted later and had generally shorter NBA careers, they did not have significantly different or reduced production than players from bigger, blue-chip conferences. Small Conference players also seemed to be assessed slightly differently by NBA front offices at the draft, as certain statistics like field goal shooting percentage and free throw shooting percentage were determinants of draft position for Small Conference players but not for Big Conference players.

Coates and Oguntimein (2010) used multiple linear regressions to assess whether NCAA production could be used to determine which performance metrics translated best to the NBA level. They performed these regressions on the entire data set of players as a

whole as well as on the Big and Small Conference groups individually. Points scored in college had a significant relationship with NBA scoring for the full sample and for Big Conference players, but not for the Small Conference sample. College rebounds and blocked shots both had a significant relationship with NBA rebounds for all of the samples, though the effect was stronger with Big Conference players. College assists, steals, and turnovers all translated well to the NBA in all samples. Free throw percentage in college, as expected, has a very strong relationship with NBA free throw percentage, though field goal shooting percentage did not.

While these results are interesting, it is important to note that Coates and Oguntimein were not technically measuring NBA production in terms of its effect on team success and winning, but rather on the accumulation of certain statistics at the professional level. More specifically, they examined which statistics transfer from the NCAA level to the NBA level when a player makes the jump. Instead of using a win-based metric like Win Shares as their dependent variable and proxy for production, Coates and Oguntimein used a variety of performance metrics like points or rebounds. This allowed them to determine if a player who scored points (or accumulated rebounds, blocks, assists, etc.) in college would do the same in the NBA, but does not determine whether any of these NCAA performance metrics predict a player's ability to help their team win at the NBA level.

In this study, Coates and Oguntimein (2010) also unintentionally provided more evidence for the difficulty of the draft and the general inability of NBA front offices to assess talent. For both Big and Small Conference players, the pick a player was drafted with did not have a statistically significant relationship with their production. In a world

where NBA teams made optimal selections in drafting players, there would be a strong relationship between draft pick number and production. Coates and Oguntimein did find a statistically significant relationship between pick number and career longevity/playing time for players who played five years or less in the NBA; however, career longevity and early career playing time does not equate to production. Moreover, this relationship with longevity is more of an indication of the stubbornness of teams to give up on high draft picks and the Sunk Cost Effect studied by Staw and Hoang (1995) mentioned previously.

Berri et al. (2011) examined which college statistics and player factors led to NBA success for players drafted between 1995-2008. Like Coates and Oguntimein (2010), Berri et al. (2011) also looked at which of these factors effected draft position. This allowed them to assess whether NBA decision-makers are using the correct college statistics and pre-draft factors when trying to identify talent and select players. College statistical data and pre-draft measurable factors of drafted players were used as the independent variables for both the draft pick and NBA production models. Players were grouped into the three traditional position groups (guards, forwards and centres) as well as by conference. Berri et al. (2011), like Coates and Oguntimein (2010), considered all players from one of the top eight conferences in NCAA basketball as Big Conference players, with the rest being grouped as Small Conference players.

The draft selection analysis used the pick number of the player drafted as the dependent variable in the regression. In terms of non-performance metrics, the results of this regression showed that both being taller and having played in the Final Four (the semi-finals of the season ending NCAA Tournament) the previous season had a significant positive impact on a player's draft stock. Of the college performance metrics,

scoring was unsurprisingly the main factor that impacted where a player was drafted. At the same time, shooting efficiency, another aspect of scoring, did not have nearly as strong of an effect on draft position. Blocks, assists, personal fouls and steals also affected draft position, though again not nearly as strongly as scoring.

The results of this 2010 study regarding scoring totals and efficiency align with other research done by Berri et al. (2007) in this area. Berri et al. (2007) determined that scoring was the most important statistic and skill for NBA front offices when assessing college players. They concluded that, if their goal was to reach the NBA, prospective NBA players should take as many shots as possible in college, even at the expense of efficiency. The inflated importance that NBA front offices place on scoring makes the trade-off of extra points for efficiency worthwhile. Berri (2006) also noted that this same points-for-efficiency trade-off makes sense for players even once they have already entered the NBA. Research on front office and coach decision making showed that general managers tended to pay higher salaries to inefficient, high volume scorers, while points per minute was the most accurate predictor for how coaches give out playing time.

For Berri et al.'s (2010) model assessing which pre-draft metrics and factors led to NBA success, the only change made to the draft selection model was the dependent variable. This dependent variable had to be a proxy for NBA performance. Berri et al. (2011) used Wins Produced per minute, NBA efficiency per minute, and Game Score per minute as these dependent variables. Instead of looking at individual, skill specific performance statistics like points or rebounds (as Coates and Oguntimein (2010) did), Berri et al. (2011) used these holistic, winning based metrics as proxies for NBA success.

Only a few player characteristics and statistics had a statistically significant relationship with future NBA production. Of the non-performance related metrics analyzed, only age had a significant effect. Berri et al. (2011) found a slight negative relationship between the age of a player when they were drafted and their future NBA production. This meant that those who played more years of NCAA basketball before entering the draft were generally less successful players in the NBA. This aligns with draft research conducted by Groothuis et al. (2007), who found that the best college players are often incentivized in a few different ways to leave college before playing out their four years of college eligibility. Players who leave college early for the NBA begin making money immediately with their cost controlled rookie contracts, reach their more lucrative second contracts earlier than those who stay in college, and can sign endorsement contracts with shoe and apparel companies as professional players. Groothuis et al. (2007) found that general managers have also become more willing to draft younger, riskier players with high upside but less refined skills. In 2004, 25 of the first 29 picks in the NBA draft were not college seniors (Groothuis et al., 2007). Furthermore, Arel and Tomas (2012) noted that between 2006-2010, 75% of first round picks were early draft entries who did not play their senior seasons. It is important to note however that age has a weaker effect on draft position in Berri et al. (2011) than in Rodenberg and Kim (2012), who examined the NBA draft age eligibility rule. This difference could perhaps be explained by the smaller draft sample (only players picked in the first round instead of both rounds) and the different timeframe (1989-2000) used by Rodenberg and Kim (2012).

In Berri et al.'s (2010) draft pick model, points scored and playing for a team in the previous year's Final Four were the only factors with a statistically significant relationship, but they had a negative effect in the NBA production model. These results are interesting in that they once again show that front offices may not be looking at the right characteristics and statistics when drafting NCAA players, as the factors that most affect draft position are not the same as those that predict future production. In the NBA production model, rebounding in college was the strongest predictor of future NBA production, and yet it was the only performance metric that did not have any sort of statistically significant relationship with draft position. Steals and 2 point shooting efficiency were also found to have a consistent statistically significant positive relationship with future production, as did assists in most of the models. In some cases, as Berri et al. (2011) highlighted, front offices based their decisions off of metrics that actually had a negative relationship with future success. Scoring was the strongest positive influencer on draft position, even though it had a negative relationship with NBA production. This once again illustrates the difficulty of the draft and the ineptitude of those evaluating talent and making draft-day decisions.

Harris and Berri (2015) conducted a nearly identical study to Berri et al. (2011), with regards to NCAA Women's basketball players and the Women's National Basketball Association (WNBA) draft. The same college statistics and pre-draft factors were gathered by the authors, with the only difference being that they considered players from any of the top 11 conferences as part of the Big Conference Group. Another important distinction between these two studies is the structure and resources of the professional leagues in question, as there is a wide gap in the potential salaries for

prospective WNBA players compared to their NBA counterparts. With significantly lower salaries available to women's players at the professional level, there is much less incentive for them to leave college early. Nearly all NCAA Women's basketball players played all four years of their college eligibility, and thus the age of a player did not have a negative relationship with production as it did in Berri et al. (2011).

Harris and Berri's (2015) draft selection model yielded similar results to Berri et al. (2011). In terms of non-performance related factors, taller players were once again drafted higher, and there was an even stronger positive relationship between draft position and playing in the Final Four in the previous season. For performance metrics, points scored had the strongest relationship with draft position. Assists and two point shooting efficiency also had a positive impact, while fouls again had a negative effect. Rebounds also notably had no effect once again, while blocks and steals did not have the same significant effect that they did in Berri et al.'s (2010) model.

Shifting to the WNBA performance model, only points, 2 point shooting percentage and fouls were significant predictors of performance, with shooting efficiency having the largest impact. Given that Harris and Berri (2015) found that where a player was drafted was not a strong predictor of their WNBA performance, WNBA front offices seem to be making similar mistakes as their NBA counterparts. While points scored in college did have some effect on the future production of prospective players, decision makers still vastly overrated this statistic compared to others. If front offices were making optimal draft picks, the pre-draft factors that effect where a player is selected and the factors that predict future performance would be one and the same.

Most recently, Ichniowski and Preston (2017) examined the effect that NCAA Tournament performance had on draft decisions and eventual NBA performance for NCAA draftees between 1997-2010. Using three different mock drafts created by scouts and media members to predict draft order, Ichniowski and Preston (2017) assessed how much a player's draft stock changed based on their personal and their team's performance in the NCAA tournament. Players who outperformed their regular season statistical output and whose teams outperformed their Tournament seeding improved their draft position relative to their average mock draft ranking; however Ichniowski and Preston (2017) found that NBA decision makers are actually undervaluing the importance of unexpected NCAA Tournament performance. These players are more likely to be successful NBA players and even become All Star players, and NBA decision makers are not accounting for these performances enough when evaluating draft eligible prospects.

3. Background

In this section, I provide a brief overview of the structure of the NCAA and amateur basketball as a whole, the importance of divisions and conferences in NCAA basketball, and the relationship between NCAA Division I men's basketball and the NBA. Before even discussing the NCAA or its relationship with the NBA, it is important to understand the history of the NBA draft as well as how NCAA Division I basketball became and remained the premier destination for amateur North American basketball players aspiring to play professionally.

3.1. History of the NBA Draft

The Basketball Association of America (BAA), which would eventually become the NBA, was founded in 1946 (Season review: 1946-47). Before this, and even during the BAA and NBA's early years, NCAA basketball was considered the highest and most prestigious level a player could reach (McCann, 2014). As the NBA became a viable career option for amateur basketball players, NCAA players began to enter the NBA draft following their college careers. NBA rules between 1946-1971 required that players spend four years in college before being declared eligible for the NBA draft or being allowed to play in the NBA (Haywood v. National Basketball Association). If they did not play in college, they were required to be at least four years removed from their high school graduation before playing. In 1971, judges in a US Supreme Court decision (Haywood v. National Basketball Association) ruled against the NBA's policy requiring players to stay four years in college if they could prove "hardship". This resulted in the drafting of Spencer Haywood to the NBA due to his financial "hardship" needs. The drafting of players who did not play in college did not become the norm however, as very few players made the jump from high school to professional basketball (Sharma, 2010). Conventional wisdom around the NBA and in basketball circles was that players needed more training and maturity before turning professional. Until the 1990's, nearly all NBA players played at least three years of NCAA basketball before declaring for the draft (Sharma, 2010).

The drafting and instant NBA success of high school draftee Kevin Garnett in 1995 marked a turning point for many NBA decision makers and thus for players with thoughts of turning professional without attending college (Sharma, 2010). From 1995-

2005, 36 high school players entered the draft (Sharma, 2010). Some of these players (Garnett, Kobe Bryant, LeBron James) enjoyed extremely successful and lucrative NBA careers, while others (Korleone Young, Leon Smith, Sebastian Telfair) barely made any impact and flamed out of the league quickly (McCann, 2004; Sharma, 2010). Regardless of whether it benefitted or hurt young players and teams, the NBA's 2005 collective bargaining agreement (CBA) stopped this practice. The 2005 CBA altered the draft to its current rules, where all players must be at least one year removed from completing their high school requirements before entering the draft (Sharma, 2010). After this rule change, nearly all top North American high school basketball players¹ have used this year to play at least one season at the college level as a means to improve their draft stock for the following year(s) (Sharma, 2010).

3.2. NCAA Division I Basketball

With all aspiring NBA basketball players now forced to be at least one year removed from high school before entering the draft, the NCAA has reaffirmed its place as the top destination for these players. It is important to note that NCAA basketball is not simply one league or group of schools, but rather a much larger, more layered structure with certain particularities. In order to understand how our data regarding NCAA players is understood and how best to analyze it, we must first examine this structure and how it affects draft eligible players and NBA decision makers.

¹ Notable exceptions to this include Brandon Jennings and Emmanuel Muddiay, who both played a year of professional basketball in Italy and China respectively before being drafted in the top 10 of the first round, as well as Thon Maker, who attended Orangeville Prep in Ontario, Canada, for a post-graduate year before being selected in the first round.

The NCAA is the main governing body for inter-university sport in the United States², and oversees 24 sports at the collegiate level. Many of these sports, including men's basketball, have multiple Divisions that represent different levels of competition. Division I is the highest level for each sport, and in the case of men's basketball, this distinction brings with it an elevated level of talent and resources (What is the NCAA). The most prestigious, historically successful NCAA men's basketball programs are found in Division I. There is also significantly more media attention and more money available to grant athletic scholarships, improve and build arenas and training facilities, and pay larger coaching and training staffs. This has created a self serving cycle where these added benefits and the heightened level of play attracts the most talented players in North America³, which simultaneously keeps the talent pool strong while increasing NCAA and school revenues. This subsequently allows schools to continue to provide scholarships and added benefits to these athletes (About NCAA Division I).

The 351 NCAA Division I men's basketball schools are also further subdivided into 32 conferences (College Men's Basketball Standings). The initial purpose of these conferences was to group teams by location in order to reduce travel times. Schools play each member of their conference at least once during their regular season conference schedule, and a conference champion is crowned at the end of the year by a single

² Simon Fraser University in British Columbia is the lone Canadian school in the NCAA, as they gained NCAA Division II status in 2009

³ Most European basketball players with professional aspirations begin playing professionally at a young age and thus are ineligible to play in the NCAA, though there has been a small recent influx of European NBA hopefuls in NCAA Division I basketball.

elimination playoff conference tournament. All conference champions gain entry into the NCAA Tournament for a chance to win the National Championship, with the rest of the field filled out by the remaining Division 1 basketball teams deemed worthy by the committee (March Madness Bracket).

With time, six conferences have emerged as the major conferences in Division I basketball: ACC, Big 12, Big Ten, Big East, Pac 10, and SEC. These conferences are consistently the six highest rated according to the NCAA's conference Rating Power Index (RPI), and they include the schools that have historically dominated Division I basketball in terms of both team and individual player success (NCAA College Basketball RPI Rankings and Ratings 2017). Just as there is a distinction between the talent level and resources available between Division I and the lower Divisions, there is a similar disparity between schools in the major conferences versus those in the others (Treme and Burrus, 2015). Schools in these other conferences are generally referred to as Small Conferences, highlighting this difference (Reinig et al., 2013). For the purposes of this study, any Small Conference is referred to as a Small Conference, while any of the six major conferences is referred to as a Big Conference. While there are undoubtedly Small Conference exceptions⁴ that have had consistent Big Conference level on-court and financial success, schools in these conferences have generally fewer resources and less talented players (Treme and Burrus, 2015). Historically, players from Big Conference schools have been drafted more often and with higher selections in the NBA draft, and are more sought after in professional leagues around the world (Berri et al., 2011). This

⁴ Recent examples include Gonzaga University, Wichita State University, and Virginia Commonwealth University.

creates another similar cyclical effect whereby young aspiring professional basketball players attend schools in Big Conferences because of the higher level of play as well as the NBA exposure and benefits they get at these schools, thereby maintaining Big Conference schools' player and resource superiority (Treme and Burrus, 2015).

While most drafted players come from the six major conferences, there are still many Small Conference players drafted. Top players can end up at Small Conference schools for a variety of reasons. Highly regarded high school players can choose one of these colleges because of a certain attachment to the school or to a coach, proximity to their hometown, or the promise of extra playing time. There are also some NBA draftees who were simply unheralded and lightly recruited high school players who only had scholarship offers to Small Conference schools and were late bloomers who improved and became professional prospects later in their college careers.

Due to the differences between major and Small Conference conferences outlined above, NBA front offices may evaluate players from these two types of conferences differently. Small Conference players play with less talented players, and thus often take on a larger role in their team's offense. They also play against generally inferior competition. Both of these factors can result in higher statistical totals for NBA draft eligible players who play in Small Conferences. Not only are these players allowed and sometimes expected to take more shots or have the ball in their hands more often, they get to do so against weaker competition than those in Big Conferences. The gap between Big and Small Conference schools is also widest in terms of the athleticism of the players, as the best high school athletes stand out to college recruiters and therefore often end up at Big Conference schools. The selection committee for the NCAA Tournament

evaluates Small Conference teams differently because of the general disparity in talent levels (Reinig et al., 2013). Playing against less athletic players at Small Conference schools can also inflate a player's statistical totals on both offense and defense, especially in terms of the number of rebounds, steals, and blocks a player can record.

NBA front offices are aware of these differences, evidenced by the fact that some of the players that record the highest statistical totals in Division I at Small Conference schools do not get drafted in the first round or even at all (Berri et al., 2011). While using percentage and rate based statistics can negate some of the impact that a team's style of play or pace can have on a player's statistics, it is much more difficult to account for the quality of a player's teammates or his opposition. This remains true even when looking at some of the newer advanced metrics available. NBA front offices may attempt to correct for the different competition that Small Conferences play against. They could even do so for players from specific Big Conferences if these conferences were perceived as weaker during the time the players played there.

4. Methodology

4.1. Theoretical Approach

In this chapter, I explain the potential factors responsible for biases and errors in decision making by NBA front offices. I also examine the sample selection bias that exists in the dataset and how to correct for it. I then describe the purpose of linear regression and the reason it is an effective technique for this analysis, as well as the assumptions that we must make in this study. Finally, I present the empirical regression models that will be used.

4.1.1. Biases and errors. In the NBA, like in the three other North American professional sport leagues, the draft remains the principle method to distribute new players onto teams. Given the effect of player quality on team quality and performance and the subsequent effect of team performance on revenue (Berri et al., 2004), the goal of NBA owners and front offices at the draft is to increase their team's chances of winning by acquiring productive players. NBA front offices therefore aim to assess incoming players and select those who will most contribute to their team's success. In many ways, this selection of players to NBA teams via the draft is akin to employers in other fields hiring new employees. With each hire, companies hope to improve the quality of their business and increase their revenue.

In typical fields of employment it is often difficult to assess and predict the future success of potential employees, especially those coming directly from university or high school with little to no prior experience in the field (Massey and Thaler, 2005). There are far fewer indicators of skill and success for firms looking to hire their next accountant or lawyer than there are for professional sports teams drafting players (Hendricks et al., 2003). In high levels of sport, player and team statistics are recorded and calculated, and there are more binary results and recordable indicators of success or failure (wins-losses, points scored, rebounds accumulated, field goal percentage, etc) from which talent evaluators can base their assessments and selections. Draftees also have experience and have accumulated statistics in the same sport that they are being drafted to play, which should theoretically make it easier to predict their future performance. This is especially true for leagues like the NBA and NFL, where nearly all drafted players have played a minimum of one season against their fellow draftees at the highest possible amateur level.

Though NBA decision makers work with all of these advantages compared to their non-sports industry peers, they still fall prey to biases and errors in their decision making. This is due in large part to a variety of psychological factors. These factors stem from decision makers' underestimation and subsequent failure to account for the uncertainty of predicting future performance. This first manifests itself in overconfidence. Massey and Thaler (2005) note that in many fields, "people believe their knowledge is more precise than it is in fact" (7). This overestimation is often augmented when more information is available. As mentioned above, NBA front offices (as well as those in other sports) are presented with more information and statistics about potential players than those hiring or recruiting employees in other fields. This extra information can actually result in even more detrimental overconfidence when it comes to predicting future performance. While this may seem counterintuitive, this is due to the fact that while added information does somewhat improve predictions, "often information increases confidence more than it increases the actual ability to forecast the future" (Massey and Thaler, 2005, 7).

NBA front offices may believe that their ability to predict the future performance of college players is greater than it really is. If this is the case, they are also likely to overestimate the importance of whatever statistical or personal factors they have deemed the most important. This type of overvaluation is an example of insufficiently regressive prediction, another issue that plagues decision makers in many fields. Insufficient or non-regressive predictions are a manifestation of the overconfidence shown by NBA front offices, as they may be prone to intuitive predictions that "are more extreme and more

varied than is justified by the evidence on which they are based” (Massey and Thaler, 2005, 6).

Most of the college basketball statistics and player characteristics that NBA front offices rely upon in making decisions are not perfectly correlated with success or failure at the NBA level. In most cases, this evidence relied upon by NBA decision makers contains a lot of statistical noise. In such instances, the importance or weight of this evidence should be reduced or at least appropriately weighted according to its reliability and correlation with success or failure. In many cases, even when decision makers are looking at the right evidence and factors when making a decision, they tend to overrate their importance, resulting in “predictions that are too extreme for all but the most diagnostic evidence” (Massey and Thaler, 2005, 6).

NBA general managers and decision makers are at risk of succumbing to this bias, as one will often hear about a team “falling in love with a player” or having their opinions about a player drastically changed based on a particular skill they noticed or even a particular game their scouts were in attendance for. While teams undoubtedly use more statistical analysis now than they used to, many of the assessments made by general managers and scouts in the NBA are intuitive. This is to say that for many teams, the decisions they make when selecting players are “reached by an informal and unstructured mode of reasoning, without the use of analytic methods or deliberate calculation” (Kahneman and Tversky, 1982, 124).

There is also the strong possibility that even when NBA teams do use statistical analysis in their pre-draft scouting, they are not emphasizing the right factors and statistics. Berri et al. (2011) noted that NBA general managers showed a strong

preference for players who simply scored a lot of points, even though this statistic actually had a negative correlation with future NBA success. They also found that the college statistics most correlated with success at the NBA level were not those that NBA front offices based their selections on during the draft (Berri et al., 2011).

One of the main barriers for overconfidence or non-regressive predictions among NBA front offices is the fact that the mere acknowledgement of these issues would be an admittance of fault or uncertainty on the part of these decision makers. As Massey and Thaler (2005) write, “to be regressive is to admit to a limited ability to differentiate the good from the great, and it is this skill that has secured NFL scouts and general managers their jobs” (7). NBA general managers and scouts are unlikely to admit that they have traditionally been overconfident with regards to their methods or place too much emphasis on certain statistics or characteristics, as it is these very abilities that they believe separate them from their peers.

4.1.2. Sample selection correction. Due to the sample of players being studied, there is also the potential for bias in our own calculations and OLS regression estimates. The goal of OLS regression is to produce unbiased estimates. In order for OLS to achieve unbiased estimates, it is imperative that the sample of data points used is an unbiased, randomly selected sample of the population being analyzed. If the sample is not randomly selected, then the OLS estimates are biased. In effect, sample selection (when data is not randomly selected) is a form of endogeneity, as we are ignoring the variable that determines the likelihood of being selected into the sample in the first place.

The NBA performance model, which tests the effect that NCAA statistics and pre-draft characteristics have on NBA performance, is susceptible to this bias. The initial

dataset for our analysis is the 372 NCAA Division 1 men's basketball players drafted to the NBA between 2006-2013. However, due to the fact that the NBA performance model is testing which NCAA performance and pre-draft characteristics effect NBA performance, only drafted players who have played enough to have their production properly assessed are included in this model. This introduces a bias⁵, as we are only selecting specific players who meet a certain game or minutes played minimum criteria while ignoring the others. The entire population of 372 drafted players is therefore not included in this particular analysis, and the sample that is used and taken from this initial 372 is not randomly selected. The draft pick model does not suffer from this sample selection bias, as this model tests the effect that NCAA performance and pre-draft characteristics have on where a player is selected in the draft, and thus all 372 drafted players are included.

To correct for the sample selection issue in the NBA performance model, it is important to apply a Heckman (1977) correction to this data. This correction is done in two stages. In the first stage, we estimate the probability of playing in the NBA given a variety of NCAA performance and pre-draft factors. This probability is then used as an explanatory variable in the NBA performance regression model. The NBA performance model will thus measure and isolate the effect that performance and player factor variables have on a player's NBA production, while accounting for the Heckman correction and the probability of drafted players playing in the NBA in the first place.

⁵ This "survivorship bias" was also discussed in Massey and Thaler's (2007) analysis of NFL draft picks and their performance

4.1.3. Linear regression. Once the Heckman (1977) correction is made, we proceed to the second stage of the analysis, where the regression models for both NBA performance and draft selection are specified. As mentioned, these regressions are used to determine the NCAA statistics and pre-draft measurable characteristics that best predict both success at the NBA level and NBA front office decision making.

This then allows us to assess the differences between what actually predicts future performance and what NBA general managers think predicts performance. Linear regression is effective when trying to estimate the effects that a multitude of independent variables have on a specific dependent variable. This statistical modelling technique is particularly useful in a social science context such as this one where the data used is not manufactured in a controlled, experimental design setting. The dependent variables in this case are influenced by a multitude of different factors that have varying levels of importance and effects. Linear regression helps to discern the individual significance of each of these variables.

In order for our OLS regression models to produce the unbiased estimates that we need, it is vital that our specifications have no missing variable bias. Missing variable bias occurs when one or more determining factors of the dependent variable are not included as independent variables in the regression specification, and these factors are correlated with one or more independent variables in the specification. When one of these variables is not included as a variable in the specification, it is consequently part of the error term. This error term is then correlated with the variables in the regression model, which violates the Gauss-Markov theorem and results in biased estimates. While the specifications for our linear regression models contain several independent variables

related to a player's NCAA performance and other factors, they obviously do not include every possible independent variable. Instead, the specifications used include statistics and factors that act as accurate proxies for the different skills and attributes that could effect a player's draft position or future NBA performance. While there are undoubtedly specific variables that are left out, every skill or attribute of importance is represented by one or more variables. This allows us to avoid any missing variable bias.

4.1.4. Assessing Decision Making Biases. Upon specifying and collecting the results of the draft and Heckman corrected NBA performance linear regression models, we can proceed to the examination of our 4th research question regarding decision making biases. In order to answer this question, we will apply general academic decision making theory, based in Kahneman and Tversky's (1973, 1977, 1979) work, to explain the disjuncture between the factors that predict draft position and those that predict future NBA performance. The statistics and factors that are either over or under emphasized by NBA decision makers could be connected to biases that plague decision makers on a more general basis. While we cannot know for certain whether or not certain biases are the cause of any faulty decision making, we will attempt to explain the results of our analyses and the decision making of NBA front offices through this aforementioned decision making theory.

4.1.5. Assumptions. In conducting this research, we assume that NBA front offices are attempting to improve their team's chances of winning through their player selections in the amateur draft. This assumption is based on the well researched idea that a winning team produces more revenue (Fort, 2011; Mongeon and Winfree, 2011; Berri and Schmidt, 2010). As Fort writes, there is a strong and positive "correlation between

revenues and winning percent” (194). NBA decision makers are thus incentivized to draft players who can help their teams win, as this will not only reflect well on their abilities as talent evaluators but will keep revenues high and their bosses (team owners) happy.

To draft the most productive players, we assume that general managers use the information that they are presented with (NCAA statistics) and that they seek out (pre-draft measurements and information) to make their selections. While it is possible that in specific circumstances, teams draft for a specific positional or player role need, most teams are simply attempting to pick the player who will be the best professional basketball player overall. Past research (detailed above) has shown that teams are generally unsuccessful at doing this, and there is very little overlap between the factors that impact NBA front office decision making and the actual factors that predict success. We are able to test whether this is still the case, and which important statistics and information are still being ignored by those who make these draft day decisions. We will be able to examine if biases exist and mistakes are being made with regards to specific player types, and if different factors predict NBA success for these player types.

4.2. Empirical Approach

As mentioned above, the model used to predict future NBA performance using NCAA statistics and data is a two-stage model. To correct for the sample selection bias, we must first construct a separate sample selection equation to estimate the likelihood of drafted players playing in the NBA. In the first stage selection equation model, a player’s likelihood of playing in the NBA is based on their NCAA performance, pre-draft physical characteristics, the pick they were selected with, and the number of years of college basketball played before being selected.

This correction term is now included in our OLS linear regression model for NBA performance. In the NBA performance based regression, an NBA career statistical proxy for a player's effect on team success is used as the dependent variable. In the NBA draft pick model, the draft pick number that a player was drafted with is used as the dependent variable. Both models use NCAA statistics and pre-draft characteristics as independent predictor variables. Each of these models is specified for the entire available data set of players as well as for each position group (Point Guards, Wings, Bigs) and conference group (Big Conference, Small Conference). More information regarding these positional and conference based groups and why they are used is provided in Chapter 5.

4.2.1. Models. The sample selection equation analyzing the NBA performance dataset is expressed as

Where H_i is a discrete variable for whether or not the player played in the NBA, expressed as 1 if the player in question played a minimum of 500 minutes in the NBA and 0 if not. X_iB_1 is a player's adjusted Win Shares per 40 minutes in college, Z_iB_1 is how many years of college basketball they played before entering the NBA draft, and T_i is the pick used to select the player.

The NBA performance (second stage) model is expressed as

Where Y is an NBA statistic used as a proxy for career NBA production across i th players, given that they have played in the NBA (determined in the stage one selection model). X_iB_1 is a matrix of NCAA performance statistics, Z_iB_2 is a matrix of pre-draft recorded physical characteristics, and C_i is a matrix of NCAA conference and position

categorical variables used as fixed effects. The variables used are explained in the next section.

The draft pick model analyzing the dataset is expressed as

Where Y is the draft pick number across i th players, $X_i B_1$ is a matrix of NCAA performance statistics, $Z_i B_2$ is a matrix of pre-draft recorded physical characteristics, and C_i is a matrix of NCAA conference and position categorical variables used as fixed effects. As explained above, this model does not require a sample selection correction due to the fact that the entire dataset of players drafted between 2006-2013 is included in this analysis.

As mentioned in section 4.1.1, overconfidence and the subsequent use of non-regressive predictions could result in biased decisions by NBA decision makers. The null hypothesis ($H_0: Y^*_i = Y_i$) is that NBA front offices show no biases related to decision making and therefore draft in an optimal way. In this case, the factors predicting a player's draft position will be the same as those that predict a player's future performance. Using the performance and draft models described above, we will test the null hypothesis by showing the varied effects of these different factors. Statistically significant factors and their effect (either positive or negative) on draft position and NBA performance could demonstrate the sub-optimal draft practices and drafting biases of NBA front offices, perhaps related to the decision making theories outlined above.

5. Data Description

In this section, I examine the reason this sample was chosen, as well as the way the data for this sample was collected and filtered. Next, I examine the variables used as well as the rationale behind the statistics chosen. Finally, I explain how players are divided into groups by their position and by the conference they played in college, and the basis for these classifications.

5.1. Data

The initial data set for this research consisted of 1071 player-seasons for each season of NCAA Division 1 men's basketball played by NCAA players drafted to the NBA between 2006-2013. This timeframe was chosen for several reasons. First, it looks at players from recent drafts, which provides information that can still be relevant to players and front offices today. Second, this research requires a large enough sample of a player's NBA career to assess their performance; including only players drafted until 2013 ensures that all players in the dataset will have had the chance to play four potential NBA seasons. Finally, 2006 was the first draft year where high school players were prohibited from entering directly into the NBA draft.

These 1071 player-seasons include each player's on court statistics for that season of college basketball, the conference their team played in, their year of eligibility during that season, and the year of that season. Each player-season also includes NBA data regarding the pick where that player was selected in the draft, their pre-draft physical measurements, and some career NBA statistics used to measure performance and longevity. All data was compiled using sports-reference.com, basketball-reference.com, foxsports.com, and draftexpress.com.

NCAA eligibility rules allow for NCAA athletes to play up to four seasons of college basketball. As mentioned, NBA draft rules also stipulate that players must be at least one year removed from high school before entering the draft. Due to these requirements, NCAA Division I basketball players could have participated in between 1-4⁶ seasons before being drafted. In order to account for this inconsistency in total seasons played amongst drafted players, only the last NCAA season played is used in this analysis. This allows each player's NBA production or draft position to be regressed along statistics from the same college sample size, while using the most recent NCAA season as a predictor. This also takes into account the information most commonly emphasized by NBA front offices when evaluating talent before the draft. The importance of a player's final pre-draft season was highlighted in Berri et al. (2011), where the authors showed that with all other factors held constant, a player's participation in the Final Four in the season immediately preceding their entry into the NBA draft raised that player's draft stock.

For the NBA performance model specifications, all NCAA players drafted between 2006-2013 who did not play at least 500 minutes in the NBA are excluded from the data sets. These players are included in the draft pick specifications though, as no NBA performance dependent variables are needed. The final data set used for analysis thus consists of NCAA and NBA statistics, pre-draft measurements, and the draft and

⁶ Robert Sacre is an exception to this, as he played 5 seasons of college basketball before being drafted by the Los Angeles Lakers with the 60th pick (2nd round) in 2012. Enes Kanter is another exception, as he signed a letter of intent to play at University of Kentucky but was declared ineligible and did not play any NCAA basketball. He is thus excluded from the dataset.

player information for all the players used in each specification. The NBA performance specifications' datasets contain the 280 players drafted from NCAA Division I between 2006-2013 who played at least 500 minutes in the NBA. The draft pick specifications' datasets contain all 372 players drafted from NCAA Division I basketball between 2006-2013. Each player's statistics and information is represented in one row of data.

5.2. Variables

5.2.1. Continuous variables used and unit of observation. In the NBA performance specifications, each player's career Relative Win Shares per Game (RWSG) in the NBA, adjusted for position played (explained below), is used as the dependent variable. Player RWSG_adj is thus the unit of observation for the NBA performance analyses. In the draft pick specifications, the pick each player was selected with in the draft is used as the dependent variable. Player draft slot is thus the unit of observations for the draft pick analyses.

The continuous independent variables used in all model specifications include NCAA Win Shares per 40 Minutes (WS40), True Shooting Percentage (TS), Effective Field Goal Percentage (EFG), Offensive Rating (ORtg), Defensive Rating (DRtg), Net Rating (NetRtg), Points per 40 Minutes (P40PTS), Rebounding Percentage (TRBPCT), Assist Percentage (ASTPCT), Steal Percentage (STLPCT), Block Percentage (BLKPCT), Turnover Percentage (TOVPCT), Usage rate (USGPCT), Three Point Rate (ThreePAR), Free Throw Rate (FTR), Height, Wingspan, Weight, and year of eligibility when drafted (Year).

5.2.2. Explanation of variables. WS40 is an estimate of the number of wins contributed to by a player according to their offensive and defensive contributions,

adjusted to 40 minutes of playing time. RWSG is an estimate of the number of wins contributed by a player per game, calculated by multiplying a player's Win Shares per minute ($WS_{40/40}$) with their NBA minutes per game.

TS is a measure of shooting efficiency that takes into account the shooting percentage of a player using the relative weight of 2 point field goals, 3 point field goals, and free throws. EFG is a measure of shooting efficiency that properly weights the importance and value of 2 point and 3 point field goals. ORtg is an estimate of the amount of points a player's team produces per 100 possessions that they are on the court. DRtg is an estimate of the amount of points a player's team allows per 100 possessions when they are on the court. NetRtg is the difference between a player's ORtg and DRtg. P40PTS is the amount of points a player scores per 40 minutes of playing time. TRBPCT is the percentage of total (offensive and defensive) available rebounds a player accumulates when he is on the court. ASTPCT is the percentage of teammate made field goals that are assisted on by a player when he is on the court. STLPCT is the percentage of opponent's possessions that end because of a steal by a player when he is on the court. BLKPCT is the percentage of opponent 2-point field goal attempts that are blocked by a player when he is on the court. TOVPCT is the amount of turnovers made by a player per 100 plays. USGPCT is the percentage of a team's plays used by a player, where their team's offensive possession ends with the player taking a shot, drawing a foul from the opposition, or turning the ball over. ThreePAR is the ratio of 3 point shot attempts to total field goal attempts taken by a player. FTR is the ratio of free throw attempts to field goal attempts taken by a player. Height is the recorded height of the player in their most recent pre draft measurement, expressed in centimetres. Wingspan is the recorded wingspan of

the player in their most recent pre draft measurement, expressed in centimetres. Weight is the is the recorded weight of the player in their most recent pre draft measurement, expressed in pounds. Finally, Year is how many seasons of college basketball a player played before being drafted to the NBA.

5.2.3. Position adjustments. For the NBA performance and draft pick specifications analyzing the entire group of players, all of the statistics and measurements used - excluding Year - are adjusted for the position of the player. From a documentation standpoint, all of these variables have a “_adj” suffix to show this adjustment (WS40_adj, TS_adj, etc.). This position adjustment is done by subtracting the position-specific average of a statistic from a player’s total in that statistic, and then adding the average of that statistic for all players back to the player’s individual total (Berri 2010). Due to the nature of the game of basketball and the different roles players are asked to play, certain player types are more likely to accumulate higher or lower numbers or percentages in specific statistics or measurements. Positional adjustments standardize these statistics and measurements across all positions.

5.2.4. Statistics Rationale. The statistics described above were chosen for a few reasons. First, when assessing a large group of players from a variety of different teams, conferences and playing styles, it is vital to remove as much statistical noise as possible. This means using statistics that isolate a player’s contributions while minimizing the effect of factors out of that player’s control. The most representative and useful statistics place players on as equal a playing field as possible. For example, using a player’s raw season totals disadvantages those whose team played less games, while even using per game statistics can hurt players who play for teams that play at a slower pace or players

that do not receive as much playing time for whatever reason. These factors are generally out of a player's control. With this in mind, the statistics chosen for this analysis are standardized as much as possible.

Using percentage-based statistics allows us to assess how much a player contributes given the opportunities they have. If two players both averaged 10 rebounds per game, and this was the only statistic we knew about the two players, it would be difficult to determine which player was better at rebounding. However, if we knew that, while they both average 10 rebounds per game, one player rebounded 30% of available missed shots when they were on the court while the other rebounded only 20%, the answer becomes clearer. When percentage stats are not available, we use other rate based statistics (i.e. P40PTS, WS40) that standardize performance across a set period of time, thereby eliminating the effect that differing minutes or games played can have on the assessment of a player's performance. Holistic measurements such as WS, NetRtg, ORtg, and DRtg also provide information regarding a player's broader impact on his team's success and on the score of the game when they are on the court, while a shooting percentage statistic like TS or EFG accounts for the relative importance of different shots and thus acts as a better indicator of how efficiently a player shoots and scores.

There are nevertheless some variables and statistics that could be important to NBA decision makers that are not accounted for in the evaluation of players. These factors may include a player's history of team and personal success at the high school, AAU and international competition levels, a player's marketability and likability, personal references from former coaches and teammates, their performance on various physiological and/or psychological tests, academic performance, family and injury

history, and team fit, amongst others. For various reasons, factors like these would be nearly impossible to include in the model or difficult to standardize for the entire group of players. For example, assessing a player's likability and acquiring a sufficient amount of personal references for each player would be extremely difficult, while there is little or no access to team and league issued physiological and psychological tests as well as players' academic performance.

5.2.5. Position groups and conference categorical variables. Discrete categorical variables are also used in this analysis. In the specifications examining the entire dataset without specific position or position-conference groups, position categorical variables are used in order to measure the fixed effect that playing a specific position has on both NBA performance and draft selection. These three categorical variables are Point Guards, Wings, and Bigs, with a player having a score of 1 in the variable of the position group they belong to and a 0 for the others.

It is important to note that the way players are grouped by position in this analysis is not by the traditional classifications given to basketball players and used in past research. The five traditional positions on a basketball team are Point Guard, Shooting Guard, Small Forward, Power Forward, and Centre. Most studies that have grouped basketball players by position have used these five position categories (Lee and Berri, 2008; Berri et al. 2010), while others have used the more broad classification of Guards, Forwards, and Centres (Sampaio et al., 2006; Abrams et al., 2008). This analysis uses only three position categories to generate larger sample sizes for each position type. However, the traditional Guards, Forwards and Centres classifications are not used.

These traditional position types were created for the way the game was played in the past, when each team's goal was to get as close to the basket as possible to take a shot. In the modern game, teams do not play with two Forwards (Small Forward and Power Forward) and one Centre on the inside; they use a maximum of two interior players and often only one. To classify those who play the small forward position and those who play the power forward position in the same group is outdated and does not reflect the way the game is now played. Small forwards play along the perimeter, while power forwards generally play on the inside. Small forwards and shooting guards have become nearly interchangeable in the modern game of basketball as "Wing" players. Power Forwards and Centres both play on the inside and have similar roles as well. While Point Guards play on the perimeter like Wing players, they play a very different role than Wings, as they handle the ball much more and are usually expected to set up and facilitate their team's offense. As Boston Celtics head coach Brad Stevens said in response to a question about the different roles of the players on the team, "it may be as simple as three positions now, where you're either a ball-handler, a wing or a big" (Goldberg, 2017). The three position types used therefore reflect this new reality, and are thus Point Guards, Wings and Bigs. Point Guards are those who play the Point Guard position, Wings are those who play the Shooting Guard or Small Forward position, and Bigs are those who play the Power Forward or Centre position. Inclusion in one of these groups is based on the position that a player was drafted to play in the NBA, rather than their listed position in college. This is thus determined by the position they have played at the NBA level, which in turn is based on the position where they have played most of their NBA seasons,

according to basketball-reference.com. These position types more accurately group together players with similar responsibilities and roles in the modern game of basketball.

5.2.6. NCAA conferences and conference categorical variables. Conference based categorical variables are also used. These variables are included to measure the fixed effects on NBA performance and draft selection of a player playing on a team in one of the six Power Conferences in NCAA Division 1 basketball. Six variables representing the six Power Conferences are used: ACC, Big 12, Big Ten, Big East, Pac 10, and SEC. Though the Pac 12 conference changed names from the Pac 10 to the Pac 12 in 2010, all players listed in either of these conferences will be considered Pac 10 players. As mentioned in Chapter 3, these six conferences were chosen because of their consistently higher Rating Power Index (RPI) scores in conference rankings as well as their traditional dominance of NCAA basketball in terms of both team and individual player success. The vast majority of both NCAA championship teams and drafted NBA players have come from these conferences. The level of talent and difficulty of playing in these conferences is therefore expected to be higher.

Players have a score of 1 in the variable of the conference they belong to and a 0 for all of the others, while a player who did not play in any of the six conferences will have a 0 in all six of these categorical variables. Playing NCAA basketball in one of these six conferences also results in a player's inclusion in the Big Conference group in the position-conference specifications. Players who did not play their NCAA basketball in one of these conferences are part of the Small Conference group.

6. Results

In this section, I present relevant summary statistics and estimates, with variations in the sample used in the models as well as what is being analyzed. There are draft pick models, performance models without a Heckman correction, and performance models with a Heckman correction. Not all estimates are examined here; only those that merit further discussion or that can be used as a potential basis for future research are mentioned. All results can be found in tables in Section 9.

6.1. Summary statistics

As stated above, the draft pick model sample contains all NCAA Division I basketball players drafted to the NBA between 2006-2013, while the NBA performance model is a selected sample containing only players who have also played a minimum of 500 minutes in the NBA with a Relative Win Share per Game score of 0 or above. The results and summary statistics presented here focus mostly on the differences and similarities between the statistics in the two models. There is a notable difference between where the average player is drafted in the full sample and where they are drafted in the selected model, as players in the NBA performance model are drafted almost six spots better (22.93) than those in the full sample (28.70). Given the requirements for the selected sample, it makes sense that this sample has a higher average draft slot. Not only should NBA teams be drafting better players who are more likely to meet this minimum minutes played threshold with higher draft selections, but even in the case that the players they select are not actually better performers, they will be more likely to grant these higher draft picks more opportunities and playing time than their lower selected counterparts, allowing them to meet the minutes played threshold.

The average NBA Relative Win Shares per Game of those in the NBA performance model sample was higher (1.91) than in the full sample (1.43). Logically, any NBA performance statistic that reflects a player's ability to affect his team's chance to win in a positive way (Win Shares, Points, Rebounds, Assists, etc) should be higher for players who have had careers or who have been given a chance to play in the NBA.

All of the other independent predictor variables and discrete fixed effects variables analyzed pertain to a player's NCAA career. This means that, unlike with the NBA draft pick and Win Share variables, the NCAA variable averages for players in the selected sample will not necessarily be higher than the averages for those in the larger full sample. Moreover, the smaller NBA performance model sample (266 vs. 372) still makes up 75% of the full sample, making it even more unlikely that the NCAA performance metrics will be drastically different in each sample. This is confirmed when examining the summary statistics. None of the NCAA performance predictor variables or the NCAA related fixed effects are notably different from one sample to the next.

With regards to the discrete categorical variables used as fixed effects, it is interesting to note that there are significantly more players drafted from Big Conferences (299) than Small Conferences (72). Moreover, Big Conference players are drafted an average of three spots higher. Of the major conferences, the ACC had the highest share of both drafted players and of players who have played at least 500 minutes in the NBA. In terms of players' positions, Wings (normally shooting guards and small forwards) and Bigs (power forwards and centers) traditionally each make up 40% of an NBA lineup, while the Point Guard position makes up the final 20% of the five man unit. The player positions of both the full draft pick sample as well as the NBA performance sample broke

down according to this traditional basketball team structure, with Wings making up 42% of the sample in each model, Bigs making up 40% and 39%, and Point Guards making up the remaining 17% and 18% in the two models.

6.2. Full Model Estimates

The draft pick model, NBA performance model without the Heckman correction, and the NBA performance model with the Heckman correction were all examined individually in order to assess and determine the importance of the different predictor variables. It is important to remember that all independent variables in these models – excluding a player's year of eligibility when drafted – are adjusted for the position played by the player. The models were then compared with each other in order to examine the similarities differences between what predicts a college player's draft position and what predicts their future NBA performance, in order to assess whether biases exist in decision making.

6.2.1. Full draft pick model. In the draft pick model (Table 9), there are several NCAA performance statistics and player factors that affect draft position. With regards to the performance metrics analyzed, higher totals in Win Shares, points per 40 minutes, assist percentage, steal percentage, block percentage and free throw rate were all statistically significant factors in raising a player's draft stock. For non-performance factors, height had a positive effect on draft position, while years played in college had a negative effect on where a player was drafted. In terms of the fixed effects looked at, playing for one of the six Big Conferences generally had a strong impact on draft position. More specifically, players from the Pac-10 conference received the biggest boost (an average of 7.93 draft positions), followed by players from the Big East (7.04),

ACC (6.31), and SEC (3.93). None of the position based fixed effects showed any statistically significant impact.

6.2.2. NBA performance models. Both NBA performance models (with and without the Heckman correction) use the same dependent and independent variables as well as the same reduced sample size. There are therefore only subtle differences between the two performance models. For the NCAA performance metrics analyzed, both models show a statistically significant positive effect for NCAA players who had higher totals in rebound percentage, assist percentage, and free throw rate. Turnover percentage was also negatively correlated with future NBA performance.

With regards to the non-performance factors analyzed, only the year of eligibility of a college player when drafted had a statistically significant impact on NBA production. In both NBA performance models, players who were younger and had less college experience were more likely to have successful NBA careers. Finally, in terms of the fixed effects analyzed, playing for a Big Conference did not have a statistically significant effect on future NBA performance, and only one NCAA conference individually had a statistically significant effect at predicting a player's future success. Drafted players from the Pac-10 conference were more likely to be more productive NBA players, while none of the other conferences had a statistically impact, be it positive or negative. Playing a specific position did not have an effect on a player's future production in the NBA.

6.2.3. Model comparison. Comparing the estimates of the independent variables in the draft pick and Heckman corrected NBA performance models, there are several similarities and differences of note. In terms of performance metrics, only assist

percentage had a statistically significant positive⁷ relationship with where a player is drafted and their future production in the NBA. The amount of years played by a player in the NCAA before being drafted had a negative relationship in both models.

On the other hand, there were several statistical and player factor variables that only had a statistically significant relationship with one of the models. Beginning with the draft pick model, NCAA Win Shares, scoring and block percentage both had a positive and statistically significant relationship with where a player was drafted. On the player factor side, relatively (to their position) taller players were also drafted higher. It is important to note that none of these factors have a statistically significant relationship with performance at the NBA level, though they have an effect on draft position.

Certain fixed effects also showed a statistically significant relationship with draft position. Playing for a team in a Big Conference, and specifically in the Pac-10, Big East, ACC, or SEC had a positive and statistically significant relationship with where a player was drafted. The Big Ten and Big 12 conference variables also had a positive relationship with draft position though it was not statistically significant. In the NBA performance model, only playing in the Pac-10 conference had a similar positive relationship. None of the other conference fixed effect variables had a statistically significant relationship with NBA performance. While their results were not statistically significant, it is interesting to

⁷ For the purposes of this study, a positive statistical relationship for the draft pick model implies that the variable is correlated with a “higher” draft selection, which will actually be lower in its numerical value (1-60). Estimates with a negative sign in the draft model will therefore be proof of a positive relationship with where a player is drafted.

note that playing for in the Big 12, Big East or Big 10 actually had a negative relationship with success, running directly counter to the results of the draft pick model.

In terms of the NBA performance model, several statistics produced statistically significant estimates as predictors of NBA success while not affecting draft position in a significant way. NCAA rebounding percentage, steal percentage and free throw rate all had a positive relationship with NBA performance for draftees. In addition, turnover percentage had a statistically significant negative relationship with NBA performance. While these statistics all acted as significant predictors of NBA success, based on the draft pick model, none of them were seen as significant by NBA decision makers.

6.3. Supplementary position based model

The first supplementary models divided players by position. Two models were specified for each position cohort, using our draft pick and NBA performance models. Both models also used the Big Conference fixed effect instead of six individual conference fixed effects. There are three position cohorts: Bigs (power forwards and centers), Wings (shooting guards or small forwards) and Point Guards (point guards). The Heckman sample selection correction was applied to the NBA performance model once again. We will examine how each of these models compares to their larger counterpart (i.e. Bigs only draft pick model compared to full dataset draft pick model), then look at how these models relate to each other (i.e. Bigs only NBA performance model vs. Wings only NBA performance model vs. PGs only NBA performance model), and finally, we will examine the similarities and differences between what affects draft position and what affects NBA performance for each positional subgroup in order to determine if any biases exist among NBA decision makers for specific position types.

6.3.1. Draft pick models. The same performance and non performance metrics that predicted draft position for the entire dataset also did so for these subsets, though there was some variation. For the NCAA performance variables among Bigs, Win Shares, points per 40 minutes, assist percentage, block percentage, and usage rate all had a statistically significant effect on draft position. Win Shares and assist percentage both had a weaker impact than in the larger draft pick model, while points per 40 minutes had a stronger one. Block percentage had a similar impact as in the larger model. Usage rate, which did not have a statistically significant relationship with draft position in the larger model, showed a significant negative relationship with draft position for Bigs.

In terms of the non-performance metrics, height, the amount of years played in college, and whether the player played in a Big Conference all had an effect on draft position. The estimates for the effect of the amount of years played in college and having played in a Big Conference on draft position for Bigs are nearly identical to the estimates in the larger model. For Bigs, taller players relative to their position group receive an even bigger boost in the draft, as the estimates for height are higher in this supplementary model compared to the larger model.

The next model contained only Wing players. In terms of the performance metrics in this model, Win Shares, rebound percentage, steal percentage; block percentage and usage rate all had a statistically significant effect on draft position. Rebound percentage and usage rate, neither of which had a statistically significant effect on draft position in the larger model, both had a significant negative impact on draft position. On the other hand, points per 40 minutes had a statistically impact on draft position in the Wings only model, though both of these factors helped predict draft position in the larger model.

For the non-performance metrics analyzed, height, years played in college, and whether the player played in a Big Conference all had an effect on draft position.

Compared to the larger model, height had a stronger positive relationship with draft position for Wings, while years played in college also has a slightly stronger negative relationship. Having played in a Big Conference has the same effect on draft position for Wings as it did for the larger dataset.

The final position based model contained only Point Guards. In terms of the performance metrics analyzed, only Win Shares, steal percentage and three-point rate had a statistically significant relationship with draft position. Three point rate – which did not have any sort of statistically significant relationship with draft position in the larger model – actually showed a negative relationship with draft position. Neither points per 40 minutes, assist percentage or block percentage had a significant relationship in this model.

As for the non-performance variables analyzed, height, weight, and years played had a significant effect on draft position among Point Guards. Height and years played in college both had nearly identical relationships with draft position in this model as they did in the larger model. Weight – which did not have any sort of statistically significant relationship in any of the previous models – had a significant and strong negative relationship with draft position for Point Guards. Finally, having played in a Big Conference – which had a positive relationship with draft position in each of the previous models – did not have any effect for the drafting of Point Guards.

6.3.2. Comparing draft pick models to each other. There are several disparities in the estimates for the different position based draft pick models. On the performance

metric side, Win Shares was the only variable that was statistically significant for all three subsets, and even then, there was a lot of variation in the effect that it had on draft position in each model; the estimates for Bigs (-3.49), Wings (-7.08), and PGs (-11.83) were all very different. Some other statistics were significant for two of the three models and some for only one of them. Predictably, block percentage had a positive relationship for Bigs (-0.031) and Wings (-0.067) but not PGs, though it was interesting that the effect was stronger for Wings than Bigs. Usage rate was also only statistically significant for Bigs (0.078) and Wings (0.068), but was actually negatively correlated with draft position.

On the non-performance side, there was slightly less impact in the effect of different variables. The number of years played in college had a negative correlation with draft position for Bigs (1.3), Wings (1.8) and PGs (1.51), while height was positively correlated with draft position for all three as well, though the effect was stronger for the Bigs (-7.29) and Wings (-7.10) than for PGs (-4.95). Playing in a Big Conference also had a positive effect on draft position for both Bigs and Wings, but interestingly no effect for PGs. Finally, weight, which has not had any statistical significance in any of the models previously specified, had a negative impact on draft position for PGs.

6.3.3. NBA performance models. The estimate for free throw rate had a similar positive relationships with NBA performance as it did in the larger model, while turnover percentage had a similar negative relationship. With regards to the non-performance metrics analyzed, the estimate for the effect of the number of years played in college for Bigs was nearly identical to the larger model.

There was not much variation between the Wings only model and the larger model as well. Interestingly, rebound percentage had no significant impact on future performance for Wings. For the non performance metrics, weight had a positive impact on future NBA performance for Wings, which had did not have a statistically significant relationship with performance in the larger model.

Finally, the Point Guards only model showed the most variation when compared to the estimates of the larger model. Win Shares had a very strong positive relationship with NBA performance for Point Guards, though it showed no statistically significant relationship in any previous performance model. The same goes for block percentage and usage rate, which had negative relationships with performance, though they showed no statistically significant relationship in the larger model. Inversely, neither rebound percentage, steal percentage or free throw rate showed a statistically significant impact for Point Guards, while all three of these statistics were positively correlated with future NBA performance for the larger dataset. Turnover percentage, normally a negative statistic, actually had a positive relationship with future NBA performance. The estimate for assist percentage is similarly interesting, as it showed a significant negative relationship with future performance, even as assist percentage was positively correlated with performance for the full dataset. In terms of the non-performance variables, none of the variables listed had a significant effect on future NBA performance for PGs.

6.3.4. Model Comparison. There are several disparities in the estimates for the different position based NBA performance models. In terms of the factors that predict future NBA performance, the Bigs only and Wings only models had several variables in

common, while the factors that predict performance for Point Guards were completely different.

For Bigs, the number of years played in college and turnover percentage were both negatively correlated with future success, while turnover percentage actually had a statistically significant positive relationship for Point Guards. Weight was a positive predictor of future performance for Wings, though it had no effect for the two other position subsets. For Point Guards, Win Shares had a positive impact on future performance, but this factor had no statistically significant impact on performance for Bigs or Wings.

6.3.3. Differences between draft and performance models. Examining the differences between the draft and performance models for Bigs, there is not much congruency between the estimates for each. The variables with similar effects on draft position and future performance are assist percentage, steal percentage, and years played in college. For the draft model, Win Shares, points per 40 minutes and block percentage were all positively correlated with draft position for Bigs, though these factors did not have a significant impact on future NBA performance. Usage rate was negatively correlated with draft position, while it had no significant effect on future NBA performance.

Switching to the Wings only model, there is even less in common with what predicts draft position and performance. There were no factors that predicted both draft position and NBA performance. Win Shares and block percentage had a statistically significant positive impact on draft position for Wings, while rebound percentage and

usage rate had negative relationships. On the other side, assist percentage and weight positively impacted future NBA performance among Wings.

Finally, the Point Guard only model showed the most disparity between the draft position and NBA performance models. Win Shares in college was the only factor that had the same type of effect for both draft position and performance. Steal percentage and height were positively correlated with draft position for Point Guards, while three point rate had a negative correlation. For the NBA performance model specifically, it is interesting to note that Win Shares had a strong positive correlation with performance for Point Guards, to the point that only one other metric, turnover percentage, had a statistically significant positive relationship with performance. Assist percentage, block percentage and usage rate all had a negative impact on future performance among Point Guards.

6.4. Supplementary Conference Based Models

6.4.1. Draft pick models. There are very few similarities between the estimates for the Big Conference and Small Conference models for both the draft pick and NBA performance models. For the draft pick models, the only common estimate is for the amount of years played in college, as there is a similar negative correlation with draft position for both models. Win Shares is the only other factor that is statistically significant for both subsets of players, though it has a much stronger effect on draft position for Big Conference players. Points per 40 minutes, assist percentage, steal percentage, block percentage, free throw rate, and height were positively correlated with draft position for Big Conference players, while weight had a statistically significant

negative impact. No other variables had any kind of statistically significant effect on draft position for Small Conference players.

6.4.2. NBA performance models. For the NBA performance models, there were no metrics that had common estimate scores between the two subsets of data. For Big Conference players, assist percentage and free throw rate both had a significant positive relationship with players' future NBA performance, while turnover percentage and years played in college had a negative impact. For Small Conference players, rebound percentage, steal percentage, three point rate and height all had a positive and statistically significant relationship with NBA performance. Interestingly, years played in college, which had a statistically significant negative effect on every other model conducted, did not have a significant impact for Small Conference players.

6.4.3. Differences between draft and performance models. For both the Big Conference and Small Conference models, there is also very little overlap between what predicts draft position and what predicts future NBA performance. Examining the Big Conference dataset, assist percentage and free throw rate were positive performance indicators of both draft position and NBA performance, while years played in college was the only non-performance metric that was negative for both the draft and performance models as well. With regards to the draft model specifically, Win Shares, points per 40 minutes, steal percentage, block percentage, free throw rate and height were all positive predictors of draft position for Big Conference players, while none of these factors helped predict future NBA performance. On the other side, turnover percentage was negatively correlated with future NBA performance for Big Conference players, though it did not have a significant effect on draft position for this subset of players.

Switching to the Small Conference models, *none* of the factors that affect draft position had any statistically significant impact on future NBA performance. In terms of predicting draft position for Small Conference players, only Win Shares (positive) and years played in college (negative) had a significant effect. Neither of these factors had an impact on Small Conference player's future NBA performance. For Small Conference players, rebound percentage, steal percentage, three point rate, and height were all positively correlated with future NBA performance, though none of these factors had any significant effect on predicting draft position for this subset of players.

7. Discussion

In this section, I discuss the relevant and noteworthy findings from the Results section. I illustrate how the draft choices made by NBA decision makers relate to general decision making bias theory and how they relate to amateur draft research from other professional sports. I then examine specific drafting errors and statistically significant variables for the general draft position and NBA performance models as well as the supplementary models in order to better understand NBA decision-making.

7.1. Drafting Errors and Decision Making Biases

Based on the results comparing the predictors of draft position and future performance in each of the models analyzed, NBA decision makers are not using the information they are presented with properly and are perhaps overconfident in their predictions for NCAA draftees. Based on their draft choices, there are several performance and non-performance factors that NBA decision makers have deemed important when evaluating college players. However, in each model analyzed, whether it

includes the complete dataset or any subset (by position or by conference), most of the factors that impact draft position did not also impact future performance. Many of these errors in decision making can be explained by general decision making bias theory, while others mirror sports biases that decision makers in other leagues have fallen prey to as well.

7.1.1. The competence hypothesis and teams' misuse of distributive data. Not only are the factors that predict draft position generally not predictive of future performance, but in every model other than the Small Conference model, there are simply more significant factors impacting draft position than performance. Though teams know that NBA drafting is volatile, they continue to bet heavily on their own expertise regarding the factors that they feel will predict which players will be successful in the NBA. This rationale can be explained by Heath and Tversky's (1991) competence hypothesis, whereby people are more likely to bet their beliefs when they feel knowledgeable about a subject. Using several test situations where subjects had varying confidence levels in their knowledge of a particular subject and question, Heath and Tversky (1991) found that whether or not this confidence is warranted, decision makers prefer to bet on decisions when they consider themselves an expert on the matter being assessed. Heath and Tversky (1991) hypothesize that this is due to the fact that, when people consider themselves experts in a subject and base their decisions off of this perceived expertise, they can thus "claim credit for a correct prediction and treat an incorrect prediction as an upset" (8).

NBA general managers and other front office members very likely consider themselves experts in their field, and demonstrate their expertise by relying on their own

ideas of what will make for a productive NBA player. This is evidenced by NBA decision makers foregoing the use of distributive data related to NCAA players drafted to the NBA, as well as their insistence on drafting players from Big Conferences who decision makers believe they know more about (explained further in section 7.3.3).

Relying on distributive statistical data could be seen as not betting on their own skills and abilities, and as Massey and Thaler (2005) write, it is these abilities that general managers often believe separates them from their peers. Moreover, Kahneman and Tversky (1977) note that while distributive data – predictive data and results that are true for a whole group on a general basis – is usually very useful, it is often ignored by decision makers. People “often express high confidence in predictions that are based on small samples of unreliable data” (24) rather than basing their decisions off larger, more robust samples of distributive data regarding the group in question. This is likely to be especially true for NBA general managers and front office members, because of the aforementioned reluctance to admit their own inadequacies in terms of prediction.

NBA decision makers foregoing the use of the distributive data and information regarding the success or failure of past draftees also results in insufficiently regressive predictions (Kahneman and Tversky, 1977). The draft decisions made by NBA decision makers are not completely intuitive in the way that Kahneman and Tversky (1977) describe, as they are evidently based in certain NCAA statistics and player factors; however, very few of these factors have been shown to actually affect future NBA performance, the outcome that they are aiming for. While decision makers in this case are not relying solely on intuition and ignoring the data they are presented with, they are not using this data properly, and are acting intuitively in *how* they use the data and what they

deem important. NBA decision makers seem to draft based on their own set of statistics and factors, instead of using empirical distributive data regarding draftees by Berri (2010) and other researchers. This is evidenced by NBA teams continuing to value some of the same pre-draft factors that have no correlation with success at the NBA level (i.e. scoring) while ignoring others (i.e. rebounding) that do. Instead of using past distributive data, NBA decision makers continue to rely on their own judgment and intuition in what factors are most likely to predict NBA success. As mentioned above, history has shown drafting in the NBA to be a difficult pursuit, and yet NBA decision makers continue to be insufficiently regressive with regards to the data they are presented with. While basing a prediction off of the data at hand is a good start, it is just as important to use the right data to make such a prediction.

7.1.2. Risk averse drafting and sticking to the status quo. Another of Kahneman and Tversky's (1979) conclusions regarding decision making is that when faced with a possible gain, people tend to be more risk averse than risk seeking in their choices. NBA decision makers seem to fall into this camp as well. Most of the experiments used in Kahneman and Tversky's (1979) analysis of risk averse or risk-seeking decision-making used situations where probabilities were known. People tended to choose the less risky options with smaller known payouts, even when their expected gain from the "riskier" option was higher, and therefore the statistically superior choice. The draft is undoubtedly an area where NBA front office members are seeking a gain, and while there are not the same kind of known probabilities in the draft as there are in the aforementioned experiments, decisions regarding who and how they draft players

seems to support the idea that NBA decision makers are applying a similar risk averse strategy.

This is evidenced again by the general tendency towards selecting far more players from the six Big Conferences, and selecting these players at higher draft positions. Upon examining all of the models, whether it is for the entire dataset or for any position specifically, players drafted after playing their NCAA basketball in a Big Conference are more likely to be drafted higher (the only exception being the Point Guard only model). However, these Big Conference players are no more likely to be successful NBA players. While it makes sense that there are slightly more players drafted from the Big Conferences due to the fact that most of the best high school prospects choose these schools, the disparity (299 to 73) should not be as great as it is given the lack of correlation between Big Conference play and future NBA performance. Moreover, smaller conference players are drafted an average of 3 spots lower than players from Big Conferences. This same preference for players from name brand colleges has been found with regards to NFL decision makers and NCAA football players (Hendricks et al., 2003), again with no statistically significant positive relationship between playing in one of these conferences and being a productive professional player.

This insistence on drafting and overrating the importance of playing in a Big Conference is likely due to the risk averse nature of decision makers. Drafting a player from a larger, more well known college from a traditionally strong conference is a safer decision. If a general manager selects a player like this and that player does not perform, the general manager is less likely to be blamed for the pick. Evidence from our analyses also show that general managers believe they are more knowledgeable about Big

Conference players and what makes these players more likely to succeed in the NBA than they are for any other subset of players. 8 different performance and non-performance factors were statistically significant in the Big Conference draft model, the most of any model. Of the 8 statistically significant factors in the Big Conference model, only three of them were significant in predicting future NBA performance. This perceived (but ultimately faulty) understanding and knowledge regarding Big Conference players can also partially explain NBA decision makers' preference for these players. In the Small Conference only model, only one factor - the amount of years played in college - had a statistically significant effect on draft position. This once again points to NBA decision makers' risk averse nature in drafting as well as the competence hypothesis at work, as they are more likely to use higher selections on players they *believe* they know more about, even if these beliefs are not grounded in fact.

Drafting Big Conference players is also an example of the status quo bias (Samuelson and Zeckhauser, 1988) at work. As Samuelson and Zeckhauser note, people have a strong tendency to keep something they already have or continue making decisions in the way that they have instead of changing or trying something different. This is due to the different way that people perceive errors of commission (choice) versus errors of omission (not choosing). In the case of NBA drafting, the status quo is to draft players from well-known colleges and conferences according to the criteria that teams have traditionally used to evaluate players. Choosing to forego this status quo and draft smaller conference players more often or use a different set of criteria to evaluate players, even if it is the statistically correct choice to make, opens up the possibility of being questioned if the selection results in a negative outcome. Sticking to the status quo and

arriving at the same negative outcome would be an error of omission that is less likely to result in questioning. NBA decision makers sticking to the status quo when evaluating college players is also another example of their risk averse behaviour when drafting.

7.1.3. Summary. In this subsection, I examined how NBA decision maker's misuse of distributive data and insistence on drafting players according to their own intuition and beliefs is explained by Heath and Tversky's (1991) competence hypothesis. Moreover, I showed how this results in a lack or misuse of distributive data, which manifests itself in insufficiently regressive predictions (Kahneman and Tversky, 1977). I then explained how NBA decision makers engage in risk averse (Kahneman and Tversky, 1979) drafting practices and suffer from a status quo bias (Samuelson and Zeckhauser, 1988) by continuing to use the same criteria for evaluating players and continuing to over-emphasize the importance of NCAA players playing in Big Conferences.

7.2. Specific Statistics in the General Models

7.2.1. The availability heuristic and the underestimated importance of ball control and efficiency. Diving deeper into the specific metrics that were statistically significant in the models analyzed, there are several interesting results that merit further discussion. Like in Berri's (2010) analysis, rebounding was once again an important skill for college players that is often undervalued by NBA decision makers. Rebound percentage for NCAA players had a statistically significant positive relationship with

future NBA performance in the general model as well as in the Small Conference model, though it had no positive effect on draft position in any model⁸.

It is important to note that in both this analysis and Berri's (2010), the proxy variable used for NBA performance is a statistic (Win Shares and Wins Produced, respectively) that values rebounding. Perhaps this can help explain why NCAA rebounding is positively correlated with higher scores in these wins based statistics. However, these individual win based statistics were chosen due to their strong and empirically tested (Berri 2006) impact on team success. This is thus likely a case where not only do NBA decision makers underestimate the importance of rebounding from the college players they will be drafting, but they underestimate the importance of this skill altogether.

Turnover percentage also had a statistically significant effect on future NBA performance but no effect on draft position. As expected, turnover percentage in college was negatively correlated with NBA performance. Turnovers are a negative statistic, as a player is ceding control of the ball and thus that possession to the opposing team. Past research regarding the importance of statistics on both an individual and team level have shown that ball control and limiting turnovers are of the utmost importance to winning in basketball (Berri, 1999; Berri and Lee, 2008; Page et al., 2007); it is therefore unsurprising that a player who has shown the ability to limit turnovers is likely to be a more productive player. Using turnover percentage instead of raw turnover totals also allows us to measure how a player limits their turnovers even when burdened with a

⁸ In the Wings only model, rebound percentage actually had a statistically significant *negative* relationship with draft position.

heavy scoring and playmaking role. This mostly negates the effect that playing style and team structure can have on a player's turnover totals, and thus can act as a better predictor of how well the player will perform in this aspect of the game in the future.

Turnover percentage and rebound percentage both relate to an important but often underappreciated skill for basketball players: ball control. Controlling possession of the ball through limiting turnovers and picking up rebounds has a demonstrated positive effect on helping teams win (Berri, 1999; Berri and Lee, 2008; Page et al., 2007), and yet both of these statistics are not valued by NBA decision makers when drafting. Limiting turnovers and second chance points from the opposition with defensive rebounding while creating more opportunities to score through offensive rebounding can increase a team's efficiency tremendously.

Similarly, free throw rate also had a positive statistically significant impact on future performance in the general model, as well as in the Bigs only, Wings only and Big Conference models, without any effect on draft position in any of these models⁹. Free throw rate is a newer advanced metric that measures a player's ability to draw fouls and accumulate free throw attempts. Modern analysis of offensive basketball has put a stronger emphasis on efficiency over raw totals, with layups, free throws and three point shots representing the most statistically efficient shots for players and teams to take. This likely explains the positive correlation between free throw rate and future success, as players who shoot a lot of free throws, and especially those who shoot free throws at a higher rate relative to lower percentage field goal attempts, help their teams score more

⁹ Free throw rate had a positive impact on draft position in the Big Conference model, but it was only significant at the 19% level and thus was not included here.

efficiently and are therefore conducive to team success. Free throw rate is also an example of the type of scoring based metric (explained in section 7.2.3.) that is based in scoring efficiency rather than raw totals. It can therefore be used as a better indicator of an NCAA scorer's likelihood of success at the NBA level rather than their rate based scoring totals, which have no correlation with future NBA performance in our models.

While solid rebounding, fundamentally sound ball control, and efficient shooting show clear links to team success, none of these statistics had a positive effect on where a player was drafted. While this seems non-sensical, perhaps it can be explained by another of Kahneman and Tversky's (1973) decision making biases: the availability heuristic. Kahneman and Tversky (1973) found that when assessing the likelihood of a particular event, people use occurrences of that event that are available for them to recall when judging its frequency. Kahneman and Tversky (1973) demonstrate that this "availability is affected by various factors which are unrelated to actual frequency" (209), and that there is "a positive correlation between the recallability of items and their apparent frequency" (222). Ichniowski and Preston (2017) discuss the availability heuristic with regards to the effect of NCAA Tournament team and individual performance on players' draft stock. Though it did not end up being the case, they hypothesized that due to this decision making bias, NBA decision makers could be overestimating the importance of the small samples of highly visible, highly memorable data from the NCAA tournament. Kahneman and Tversky (1973) explain that factors that "strengthen the association between item and list should increase the apparent frequency of the item" (221). The frequency or importance of specific events or "items" can be overestimated when they are more memorable and available for recall.

In the case of player evaluation, Kahneman and Tversky's "items" are basketball plays, events, or statistics, while the "list" is positive characteristics and skills that lead to success. Highlight reel dunks, blocked shots and other feats of athleticism, along with high point totals and frequent scoring, are undoubtedly eye catching and memorable. For most people, they are more memorable than rebounding, limiting turnovers, and playing the game in an efficient way. The overemphasis of these more noticeable and available events at the expense of comparably less flashy (but perhaps more important) skills could therefore be a case of the availability heuristic at work. Due to the memorable nature of the aforementioned dunks, blocks, scoring and athleticism, NBA decision makers are likely to recall them happening more frequently than they do in reality. Moreover, it is simple to see how these events are positive for their team's success, strengthening the association between these types of plays and a player's success. As Kahneman and Tversky write, "in evaluating the probability of complex events (such as the effect of a particular statistic or skill on team success)¹⁰, only the simplest and most available scenarios are likely to be considered" (229).

Lending more credence to this idea of the availability heuristic playing a part in NBA teams' decision making is the fact that both scoring and blocked shots have a statistically significant positive impact on draft position, with no effect in predicting future NBA performance. It is easy to see how scoring a high number of points or blocking an opponent's shot would have a positive impact on team success. However, perhaps these points come at low efficiency, or the blocks do not truly occur at a high

¹⁰ Parentheses added by author

enough rate to make a tangible, consistent difference in games. In reality, the ability to control the ball and play efficiently is what helps win games, and hence what predicts how a college player will help his team win games in the NBA. However, neither rebounding, efficient shooting, nor ball control is very memorable, with the latter two factors manifested in the *lack* of a particular event (turnovers or missed shots) rather than in actually completing a specific task or recording a certain statistic. Skills of omission like these are thus likely even less memorable, and players with these skills, though empirically valuable, are even more likely to fall prey to the availability heuristic when assessed by NBA decision makers. While highlight reel plays as well as imposing physical stature (which will be discussed in further detail in section 7.2.2) may draw more of the attention, it is likely the simpler skills that actually predict how a player will perform in the NBA.

7.2.2. Age, size and the importance of potential. The amount of years played in college was the most consistent metric across all of the models, as it had a negative impact on both draft position and NBA performance for every model except for the Wing only, Point Guard only, and Small Conference performance models. As mentioned previously, because of the nature of NBA draft rules, the best NBA prospects are forced to be at least one year removed from high school before entering the draft, and will therefore often play only one or two years of college basketball (Groothuis et al., 2007; Arel and Tomas, 2012). This likely explains why there is a negative correlation between the number of years played in college and both draft position and success; it is not that younger players are better suited to play in the NBA, but rather that the best prospects leave college earlier. While this was true for nearly all subsets of data, it is interesting to note that one

of the groups that it did not impact was Small Conference players. This is likely due to the fact that many of the Small Conference players who reach a level to be considered for the NBA draft are late bloomers, which is the very reason they likely went to a college in a smaller conference to begin with.

The strong negative correlation between the number of years played in college by NBA draftees and draft position also highlights NBA decision makers' desire to draft for potential. It is often said that NBA general managers look for potential as much as past performance when drafting, and these results align well with this idea. Even excluding the top players, who likely would have foregone college altogether and entered the draft straight from high school if NBA rules did not prohibit it, NBA teams are still often interested in younger, more raw players. As an NBA assistant coach told reporter Kevin O'Connor, "The draft is all about upside, so teams make the same mistake every time. [They'd rather take projects like] OG Anunoby than a seasoned player like Kuzma," (O'Connor, 2017). NBA decision makers are more likely to take a flyer on a young player with prototypical NBA size and athleticism, even if they have not demonstrated high-level basketball skills or production on the court. The same search for potential seems to be present in draft research in MLB. MLB decision makers have traditionally overvalued high school players relative to more experienced, older college players, even though college players have been shown to be consistently better and safer choices (Spurr, 2000; Burger and Walters, 2009). MLB teams often marvel at the potential of younger high school players, who likely post incredible statistics against diluted and vastly inferior competition in high school baseball. While some of these high school players do turn out to be stars at the MLB level, their abilities and performance against

other high school players is much less representative of how they will perform at the professional level compared with college players, who are playing against physically mature peers with much more comparable skills and ability. As Hurley et al. (2012) found with regards to hockey players, picking younger players often brings with it extra risk, as it is harder to project a player's ability when they are younger, playing against inferior competition, and there is less data on which to base their performance. However, like in the NBA, decision makers in other sports often find the potential of a young, athletic player who looks the part to be too tempting to pass up.

The results regarding the importance of the height and weight of draftees reinforce this desire to draft for potential. In the general model, both height and weight were positively correlated with higher draft position, yet neither of these factors had a statistically significant relationship with future NBA performance, except for weight with regards to Wings.

NBA decision makers seem to be over emphasizing the importance of size when evaluating players. Instead of focusing on the abilities of a player and how they have actually fared on a basketball court in terms of helping their teams win, NBA decision makers look at the potential of a player given their size and athleticism and what they can be turned into at the next level. They believe that they can teach and develop the other skills and basketball savvy necessary to succeed, but no amount of coaching or instruction can change a player's body; as famed Boston Celtics' coach Red Auerbach said, "you can't teach height".

This coincides with draft and decision-making research done in other professional sports as well, where both size and potential are often over emphasized. In Voyer and

Wright's (1998) analysis of Canadian Junior Hockey players drafted to the NHL, they found that goals and assists in Junior in both the regular season and playoffs were the only correlates with success at the NHL level. Size (height and weight) had no effect on a player's likelihood of success, even though this has traditionally been used as a tool to evaluate the potential of players. NHL teams and decision makers still often talk about the need to get bigger and draft and acquire taller, heavier and stronger players, even though time and time again, research has shown size to have no correlation with production. Much of this belief stems from NHL decision makers' outdated ideas of what player attributes relate to production in hockey. At the same time, this desire to acquire bigger players also relates to general managers' fascination with potential, as they often believe that if a player has the requisite size and build of a star player, the other on ice or on court skills can be taught and the player can develop into a productive member of the team.

7.2.3. A continued over-emphasis on scoring. Another interesting conclusion from the models has to do with scoring at the college level. Berri (2010) found that scoring was the most overrated factor for NBA decision makers, as it had a strong positive impact on where a player was drafted but was actually negatively correlated with NBA performance. Our analyses yielded similar results, as scoring had a positive impact on draft position for the general model and an even stronger positive impact for the Bigs only and Big Conference models, while having no significant impact on performance in any of the models. Previous work by Berri and Schmidt (2010) regarding the determining factors of team success as well as employment and salary in the NBA illustrate this same

over-emphasis on scoring, though this has not deterred NBA decision makers from continuing to make similar decisions.

This is not to say that scoring is unimportant or that players who score at higher rates are necessarily unproductive players, but rather that there are other factors that are important when projecting how NCAA basketball players will fare in the NBA. There are of course many NCAA players who score at high levels in college and then go on to become stars in the NBA, but these players were likely successful in other areas of the game as well. This once again relates back to the status quo bias of NBA decision makers, as the selection of a high scoring player – even if they do little else of value for their team on the court or are even detrimental in certain areas – will rarely if ever be met with questioning. On the other hand, selecting a player who's signature skill is his ability to shoot efficiently or rebound – factors that have a positive impact on NBA performance in both Berri's model and in those above – would likely be seen as more of a risky or strange pick to those in the basketball world.

7.2.4. Summary

In this subsection, I looked at data regarding specific variables in both the draft and performance models. First, I discussed the oft-ignored importance of ball control and efficiency when drafting NCAA players, based on results regarding the link between rebounding, turnovers, free throw rate and future NBA success. This subsection also examined how the availability heuristic (Kahneman and Tversky, 1973) likely explains the reason NBA decision makers continue to ignore these factors, instead focusing on more eye catching abilities and statistics. Next, I showed the way the results regarding the perceived importance of player age and size demonstrate NBA decision makers'

insistence on drafting for potential, and the way this aligns with research regarding other professional sports leagues as well. Finally, I examined NBA decision makers' continued over-emphasis on the importance of scoring for NCAA draftees, which mirrors Berri's (2010) research.

7.3. Supplementary Model Specific Biases

7.3.1. Bigs only and Wings only model. Some of the results and differences in the supplementary models between what predicts draft position and performance also merit further discussion. Beginning with the Bigs only model, it is interesting to note that usage rate had a statistically significant negative relationship with draft position, but no effect on future NBA performance. The same was true in the Wings only mode, as Wing players with a higher usage rate in college actually saw a negative effect in their draft position. This could once again be related to NBA decision makers' insistence on drafting for potential; many of the physically impressive, athletic players whose skills are raw and underdeveloped do not play a featured role in their team's offense. This is especially true with athletic, "project" Bigs, whose lack of fundamental skills to match their athleticism often result in them playing much more complementary offensive roles in college. In these cases, they are often simply catching passes near the rim after a play is made by someone else or scoring off of offensive rebounds and hustle plays. Players with these more developed skills are often older and do not have the eye popping size or athleticism of the younger players who get drafted higher. While it is true that some of these raw athletic marvels end up as productive NBA players and stars, many also flame out of the league; the reverse is true as well, as some older, more polished players simply do not have the size or athleticism to play in the NBA, but many are able to transfer their

skills to the next level and become productive NBA players. This can perhaps explain how a statistic like usage rate could be a negative for a player's draft position.

As mentioned above, free throw rate was also a predictor of performance for Bigs, but had no effect on draft position. This is interesting given the fact that this is a statistic that is related to the work players do closer to the basket. For Bigs, getting to the free throw line is often a result of the effort these players are willing to put forth. Players who do this at the NCAA level are often labeled as hustle players, a sort of backhanded compliment that lauds their hard work while also subtly commenting on their perceived lack of skill or athleticism. The results of this analysis highlight the importance of effort, and also show that perhaps effort and relentless play is a skill in its own right that should be treated as such by NBA decision makers.

7.3.2. Point Guard only model. Switching to the Point Guards only model, Win Shares had by far the strongest statistically significant positive impact on future performance, to the point that the other significant performance based estimate scores in this model had a negative effect on performance¹¹. The estimate for the effect of Win Shares on performance (0.75) was nearly 10 times higher than the next highest score in any model (Big Conference model - 0.083). This is a very interesting finding that makes a lot of sense given the traditional role of a Point Guard. Point Guards are often asked to be the leaders of their teams, in charge of running a team's offense and setting their teammates up, getting other players in the correct spots on offense and defense, and

¹¹ The only exception to this is turnover percentage, which also had a negative sign but, as discussed above, this shows that limiting turnovers was once again an important factor for predicting NBA performance.

generally being an extension of the coach on the court. Coaches will often talk about their Point Guard's intangible leadership and impossible to quantify skills that cannot be measured on the usual stat sheet by points, rebounds or assists.

With this in mind, it makes sense that a statistic like Win Shares, which measures a player's overall effect in their team's success or failure, would be strongly correlated with a Point Guard's future NBA success. With the dependent proxy variable for NBA performance used in our analysis being NBA Win Shares, it would seem likely that NCAA Win Shares would be a strong predictor of future performance for all players; however, this was not the case in the other models analyzed. Only the Point Guard model demonstrated this strong relationship, perhaps because this is the position where a player's overall contribution to team success is the most important. Rather than any particular individual statistic, it is a Point Guard's ability to lead their team to wins – the same ability that is most important when evaluating a coach – that is the best predictor of their future individual success in the NBA.

7.3.3. Big and Small Conference only models. Examining the Big Conference model, it is important to note that is subset has the largest sample of any of the supplementary models, counting 299 of the 372 total players. The results are therefore very similar to the general model, with some notable differences. As mentioned above, there are 10 different statistically significant predictors of draft position in the Big Conference model, two more than the next closest model (the general model has eight). Steal percentage and turnover percentage were the two extra statistically significant factors in the Big Conference model, to go along with the same eight as the general model. While the Big Conference draft model has the most predictors, only three of these

10 factors actually had an effect on NBA performance. This once again points to NBA decision makers emphasizing the wrong statistics and player factors when evaluating NCAA players, even when they belong to a cohort (Big Conference players) who decision makers believe they know the most about.

Interestingly, the amount of statistically significant factors for draft position and NBA performance is the reverse for Small Conference players. Only two variables had any sort of significant impact on draft position for these players, the lowest total for any draft model. For the Small Conference NBA performance model, four different factors acted as significant predictors of NBA performance, none of which had any sort of impact on draft position. The lack of predictive factors for the draft position of Small Conference players is in sharp contrast to the high number for Big Conference players, lending more credibility to the idea that NBA general managers believe that they know much more about Big Conference players. This subsequently supports the aforementioned idea that NBA decision makers, like the general population observed by Heath and Tversky (1991), are more likely to bet on a subject that they think they know more about. Smaller conference players are selected much less often and with lower picks, and while some of this has to do with the better high school prospects choosing Big Conference schools, the results point to the competence hypothesis having an effect as well. This competence hypothesis bias, coupled with the status quo bias mentioned earlier, can explain the bias that NBA general managers have towards the selection of Big Conference over Small Conference players.

7.3.4. Summary. In this final section, I examined the results of the supplementary position and conference specific models. In the Bigs only model, I illustrated how the

polarized effect of usage rate in the draft and performance models is perhaps another reflection of NBA decision makers' insistence on drafting for potential. I also discussed how the importance of rebounding and free throw rate for Bigs points to effort being a skill that NBA decision makers should look at when evaluating draftees. Next, in examining the Point Guard only model, I explained how the extreme importance of Win Shares above all other statistics makes perfect sense given the role of a Point Guard as the leader and coach-on-the-floor of their team. Finally, with regards to the Big and Small Conference models, I examined the large disparity in predictive factors for the draft position model in the Big Conference (10) and Small Conference (1) models, and how this disparity aligns with results indicating that Big Conference players are drafted more often and with higher selections. These results offer another example of the competence hypothesis at work, as NBA decision makers prefer Big Conference players who they believe they know more about, even though these players are no more likely to be productive NBA players.

8. Conclusion

In this paper, I examined decision making in the NBA through an analysis of the factors that predicted draft position and those that predicted future performance for NCAA Division I basketball players drafted to the NBA between 2006-2013. While NBA draft research has been conducted in the past, others have used older, outdated statistics, metrics, and position groups for players both at the college and NBA levels, and no previous research has corrected for the sample selection issue inherent to research regarding the future performance of draftees. These changes allow for more accurate and

isolated evaluation of players as well as empirically stronger regression specifications. The results of these specifications are then linked to several general decision making theories and biases.

Upon examination of the similarities and differences between the draft position and performance models, I was able to infer several interesting conclusions not only regarding individual statistics and metrics, but what these results signify for the way NBA front office members make decisions generally. While it is interesting to understand the specific statistics and factors that NBA decision makers overrate (Win Shares, points, blocks, height, and Conference) as well as those that they mistakenly ignore (rebounding, steals, turnovers, free throw rate), there is just as much if not more value in understanding the general decision making biases and theory that affect their decision making processes.

NBA decision makers rely on their own expertise rather than on distributive data when making draft choices, a mistake explained by Heath and Tversky's (1991) competence hypothesis. They also have a tendency to stick to the status quo (Samuelson and Zeckhauser, 1988) and use generally risk averse strategies in their decision making. NBA decision makers also seem to fall prey to the availability heuristic when simultaneously undervaluing the importance of certain more subtle statistics like rebounding and turnover percentage while amplifying the value of others. Finally, like their counterparts in other professional North American sports leagues, NBA decision makers overemphasize the importance of potential at the expense of known value.

Future research in this area could move forward in two different directions. With the data at hand and the availability of data of this kind, one can conduct a variety of more specific draft analyses. Many drafted players also come from a select number of

schools, which could lead to an interesting study. Specific analyses based on the amount of years a player plays in college before entering the draft could also yield significant results regarding the differences between younger and older draft eligible players. There are many ways to parse out this specific dataset or to examine a different set in ways that could improve our understanding of what factors affect a college player's drafting and transition to the NBA. Looking at a different pool of data, one could also conduct a similar analysis on undrafted players who have played a significant amount of time in the NBA, in order to determine what factors predict success for these players. It would be interesting to see whether successful undrafted players had similar statistics and characteristics as their drafted counterparts.

On the other hand, this research could open the door to further examination of decision making in the sport context. As illustrated above, many of the same decision making theories, biases and errors in judgement that apply in a general context also affect NBA decision makers. Given the similar draft structure and general goals of professional sports organizations, there is a possibility that many of these decision making theories apply to other leagues and decision makers as well. Similar studies can be conducted examining other professional sport league drafts, as well as more general coaching and front office decisions in the NBA and other leagues.

History has shown that drafting in any professional sport is difficult. It is difficult to project how players who have never performed at the highest professional level will perform when they get there, even as more information becomes available. Simply being aware of their own human biases could be the first step in improving drafting by decision makers. Even NBA general managers, scouts, and front office members, for all of their

perceived and real basketball expertise, are human. They suffer from biases and make errors in judgement that many – in analogous situations in nearly any decision making context – would make as well. This does not mean NBA decision makers must drastically change their evaluation, scouting, or the way they draft, but rather that they should perhaps remember that they are only people, and must account for their own decision making biases just like anyone else.

9. Tables

Table 1. Summary Statistics

	Summary Statistics for Full Dataset			
	Mean	SD	Min.	Max.
<i>Dependent variables</i>				
Draft slot	28.7	17.15	1	60
Log draft slot	3.13	0.82	0.69	4.11
NBA relative win shares per game	0.029	0.033	-0.032	0.175
Log NBA relative win shares per game	0.028	0.031	-0.033	0.16
<i>Independent variables</i>				
Win Shares per 40 minutes	0.2	0.051	0.1	0.42
Log Win Shares per 40 minutes	0.18	0.04	0.09	0.35
Points scored per 40 minutes	19.78	4.02	9.83	33.96
Log points scored per 40 minutes	3.01	0.19	2.38	3.55
Rebounding percentage	11.73	3	2.38	21.81
Assist percentage	15.59	5.99	2.69	44.08
Steal percentage	2.2	0.79	0.65	4.75
Block percentage	3.6	2.39	-2.44	16.05
Turnover percentage	14.57	3.42	6.69	25.93
Usage rate	25.09	4.49	13.05	38.51
Effective field goal percentage	54.42	4.63	42.42	71.72
Three point attempt rate	23.47	12.96	-9.39	71.2
Free throw rate	43.71	13.33	6.52	85.4
Height (cm)	201	4.53	187.03	214.03
Log height	5.3	0.022	5.23	5.37
Weight (lbs)	217.95	16.63	169.92	280.92
Weight	5.38	0.075	5.14	5.64
Year of college eligibility when drafted	2.86	1.1	1	5
Log year of college eligibility when drafted	1.3	0.32	0.69	1.79

*Fixed effect**variables, equal to**one if...*

Player played in a Big Conference	0.8	0.39	0	1
Player played in a Small Conference	0.19	0.39	0	1
Player played in Atlantic Coast Conference	0.16	0.36	0	1
Player played in Big 12 conference	0.13	0.34	0	1
Player played in Big East conference	0.15	0.36	0	1
Player played in Big Ten conference	0.08	0.27	0	1
Player played in Pacific 10 (or 12) conference	0.13	0.33	0	1
Player played in Southeastern conference	0.13	0.34	0	1
Player has played the point guard position in the NBA	0.17	0.38	0	1
Player has played the wing position in the NBA	0.42	0.49	0	1
Player has played the big or post position in the NBA	0.4	0.49	0	1
Player was drafted in the 2006 draft	0.11	0.32	0	1
Player was drafted in the 2007 draft	0.1	0.31	0	1
Player was drafted in the 2008 draft	0.12	0.33	0	1
Player was drafted in the 2009 draft	0.12	0.33	0	1
Player was drafted in the 2010 draft	0.14	0.35	0	1
Player was drafted in the 2011 draft	0.12	0.32	0	1
Player was drafted in the 2012 draft	0.13	0.34	0	1

Player was drafted
in the 2013 draft

0.12

0.32

0

1

Observations: 372

Table 2. Summary Statistics

Summary Statistics for Heckman Corrected Sample				
Variables	Mean	SD	Min.	Max.
<i>Dependent variables</i>				
Draft slot	22.93	14.96	1	60
Log draft slot	2.9	0.81	0.69	4.11
NBA Relative win shares per game	0.039	0.031	-0.012	0.17
Log relative win shares per game	0.04	0.029	0.001	0.16
<i>Independent variables</i>				
Win shares per 40 minutes	0.21	0.05	0.1	0.42
Log win shares per 40 minutes	0.19	0.04	0.1	0.35
Points scored per 40 minutes	20.08	4.17	9.83	33.96
Log points scored per 40 minutes	3.02	0.19	2.38	3.55
Rebounding percentage	11.92	3.03	2.38	21.68
Assist percentage	15.71	5.96	2.79	44.08
Steal percentage	2.27	0.81	0.65	4.75
Block percentage	3.71	2.39	-2.44	16.05
Turnover percentage	14.47	3.4	6.69	25.93
Usage rate	25.36	4.54	13.05	38.51
Effective field goal percentage	54.42	4.41	44.5	66.8
Three point attempt rate	22.46	11.7	-9.19	61.4
Free throw rate	44.2	12.91	11.12	79.32
Height (cm)	201.3	4.54	190.6	214.03
Log height	5.3	0.022	5.25	5.37
Weight (lbs)	218.58	16.3	169.92	275.92
Log weight	5.38	0.075	5.14	5.62
Year of college eligibility when drafted	2.67	1.1	1	5
Log year of college eligibility when drafted	1.25	0.32	0.69	1.79
<i>Fixed effect variables, equal to one if...</i>				

Player played in a Big Conference	0.81	0.39	0	1
Player played in a Small Conference	0.18	0.39	0	1
Player played in Atlantic Coast Conference	0.16	0.37	0	1
Player played in Big 12 conference	0.12	0.32	0	1
Player played in Big East conference	0.15	0.36	0	1
Player played in Big Ten conference	0.08	0.28	0	1
Player played in Pacific 10 (or 12) conference	0.14	0.35	0	1
Player played in Southeastern conference	0.13	0.34	0	1
Player has played the point guard position in the NBA	0.18	0.38	0	1
Player has played the wing position in the NBA	0.42	0.49	0	1
Player has played the big or post position in the NBA	0.39	0.48	0	1
Player was drafted in the 2006 draft	0.13	0.34	0	1
Player was drafted in the 2007 draft	0.09	0.29	0	1
Player was drafted in the 2008 draft	0.11	0.31	0	1
Player was drafted in the 2009 draft	0.14	0.35	0	1
Player was drafted in the 2010 draft	0.12	0.32	0	1
Player was drafted in the 2011 draft	0.12	0.33	0	1
Player was drafted in the 2012 draft	0.14	0.35	0	1
Player was drafted in the 2013 draft	0.11	0.31	0	1

Observations: 280

Table 3. Parsimonious Models

Parsimonious Test Models				
Model	(1) Performance Model	(2) Performance Model	(3) Performance Model	(4) Draft Pick Model
Win Shares	0.21* (0.034)	0.20* (0.030)	0.20* (0.0429)	-7.90* (0.79)
Height	0.00032 (0.00038)	0.00060 (0.00033)	0.056 (0.072)	-3.71* (1.47)
Year	-0.0079* (0.0016)	-0.010* (0.0013)	-0.0067* (0.0017)	1.33* (0.1)
Heckman Sample	N	Y	Y	N
Heckman Correction	N	N	Y	N
Observations	280	266	266	372

*** ** * indicates statistical significance at the 10%, 5%, and 1% levels, respectively. The standard errors are in parentheses.

Table 4. Heckman Sample Selection First Stage Model*Heckman as dependent variable*

Pick	-0.045*
	<i>0.010</i>
Win Shares	7.88*
	<i>3.11</i>
Year	-0.39
	<i>0.15</i>
Observations	266

*** ** * indicates statistical significance at the 10%, 5%, and 1% levels, respectively.

The standard errors are in parentheses.

Table 5. Draft Pick and Performance Models

Models	(1) Draft Pick Model	(2) Draft Pick Model	(3) Performance Model	(4) Performance Model
Full sample of all positions and conference				
Win Shares	-5.23* (1.32)	-5.4832*	0.064 (0.069)	0.0639 (0.0700)
Points per 40	-0.80*** (0.55)	-0.7704*** (0.5590)	0.0077 (0.026)	0.0095 (0.0269)
Rebound Percentage	0.012 (0.01)	0.0057 (0.0145)	0.0012*** (0.00064)	0.0012** (0.0007)
Assist Percentage	-0.21* (0.007)	-0.0231* (0.0079)	0.00079** (0.00035)	0.0007* (0.0003)
Steal Percentage	-0.062*** (0.047)	-0.0645 (0.0486)	0.0037*** (0.0023)	0.0036*** (0.0023)
Block Percentage	-0.029*** (0.07)	-0.0383** (0.0163)	0.000063 (0.00080)	0.0002 (0.0008)
Turnover Percentage	-0.0007 (0.013)	-0.0012 (0.0130)	-0.0015** (0.00066)	-0.0014** (0.0007)
Usage Rate	0.0136 (0.023)	0.0072 (0.0032)	-0.00063 (0.0010)	-0.0006 (0.0010)
Effective Field Goal Percentage	-0.0059 (0.011)	-0.0052 (0.0117)	0.00029 (0.00054)	0.0003 (0.0005)
Three Point Rate	0.0017 (0.003)	0.0007 (0.0033)	0.00014 (0.00016)	0.0001 (0.0002)

Free Throw Rate	-0.0046*** (0.003)	-0.0043*** (0.0031)	0.00036* (0.00014)	0.0004* (0.0001)
Height	-5.45* (1.76)	-5.1767* (1.7460)	0.083 (0.085)	0.0837 (0.0858)
Weight	0.34 (0.47)	0.4809 (0.4798)	0.015 (0.023)	0.0141 (0.0232)
Year	0.38* (0.031)	1.2828* (0.1765)	-0.025* (0.0058)	-0.0245* (0.0058)

Fixed Effect Variables

Big Conference		-0.2750* (0.089)		0.0026 (0.0044)
Bigs		-0.0056 (0.071)		0.0021 (0.0034)
Point Guard		0.0662 (0.091)		0.0043 (0.0046)

Heckman Correction	N	N	y	Y
Observations	372	372	266	266

*** ** * indicates statistical significance at the 10%, 5%, and 1% levels, respectively. The standard errors are in parentheses.