An Empirical Analysis of the NBA Draft from 2006-2014

Lloyd Ellison

Abstract:

This paper investigates how well the NBA (National Basketball Association) does at drafting talent that succeeds. There has always been a question of if NBA prospects can have their NBA talent forecasted. The study seeks to compare different draft years and compare if the model can determine which athletes will become successful. The study seeks to use college statistics as well as the NBA draft combine data (which includes height, weight, wingspan, etc.) The study will be cross-sectional to get the best view of a bunch of players and because depth of talent can vary year to year.

JEL Classification: Z20, Z21, Z22
Keywords: Sports Economics, NBA.

[a] Department of Economics, Bryant University, 1150 Douglas Pike, Smithfield, RI02917.
Phone: (617) 999 7362. Email: lellison@bryant.edu.

1.0 INTRODUCTION

Drafts in sports are a unique thing to North America and due to the nature of the National Basketball Association (NBA) and the impact that one player can have on a team. This paper aim to enhance the understanding of what attributes make a successful NBA player and how well does the NBA do at drafting top talent.

The first major league in America to create a reverse order draft was the NFL in 1936. The idea behind a reverse order draft is to create competitive balance and this true in all leagues. Different leagues have different rules behind the draft. For instance, in the NBA the top 14 draft picks are decided through a lottery in which if a team has done worse, they have a better chance at getting a top pick, but not a guarantee. Both in the NBA and NFL draft picks are allowed to be traded which inherently puts value in them. The MLB does not allow draft picks to be traded but does allow teams to receive a compensation draft pick from a team that signs a player that the original team offered a qualifying offer. The goal of this is that franchise players have incentives to stay and if they do not, then the team gains an advantage. The MLB also has a competitive balance round which picks randomly 10 teams from the smallest markets and lowest revenue clubs to give a better chance at drafting great talent. The length of drafts also changes from league to league. The NBA draft is only two round each consisting of 30 picks. The NFL has 7 rounds of 32 picks. The MLB has 40 rounds consisting 30 picks (plus compensation and competitive balance). Finally, The NHL has seven rounds of 31 picks.

Like most sports the NBA has had multiple drafts where top picks are used on players who don't last very long in the league sometimes almost immediately failing and other times just sputtering out. So why do players like Anthony Bennett get drafted number one overall? Did scouts or general managers miss something that was in the data? Did Bennett just struggle on his own accord? This paper aims to see if there is something missed in the data. On the flip side are player like Jimmy Butler who was drafted 30th but has had a long successful career. One difference is that Jimmy Butler slowly got better but could this be projected. Not every star starts out great, but there is a reason that they are drafted towards the top. More recent examples have been first overall pick Markelle Fultz who did not perform well with the 76ers (some of which was due to injury) and then was traded to the Magic. Although better with the magic it is not clear if Fultz can perform to his original lofty expectations as a number one pick overall.

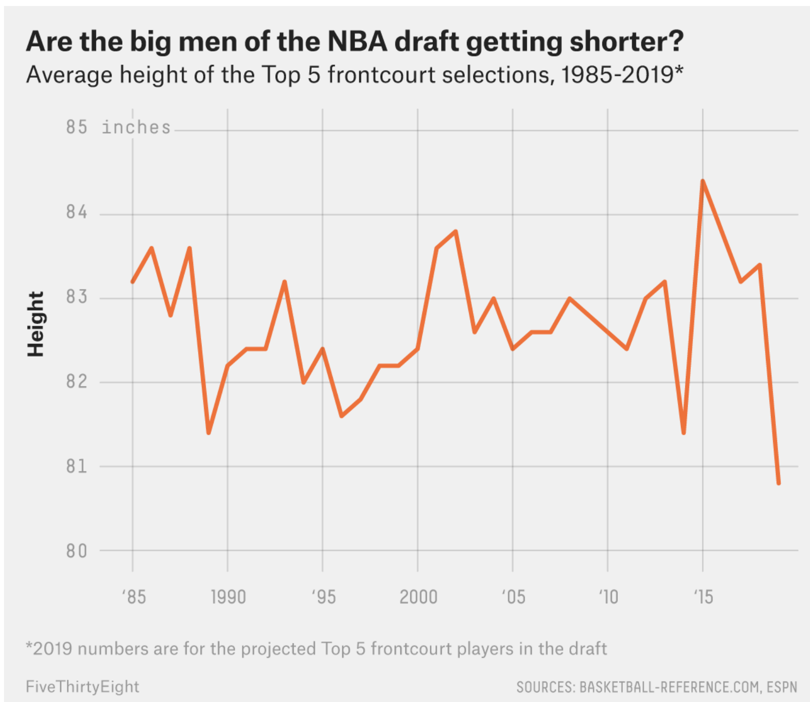
This paper was guided by three research objectives that differ from other studies: First it investigates the NBA using combine data as well as college basketball data; Second, it incorporates several different models including classic econometric techniques as well as new prediction models; Last, it analyzes the NBA draft and success using several different metrics including games played, years played, if given an award (NBA first tea), and how many awards they have been given. This paper successfully fills this void.

The rest of the paper is organized as follows: Section 2 gives a brief literature review. Section 3 outlines the empirical model. Data and estimation methodology are discussed in section 4. Finally, section 5 presents and discusses the empirical results. This is followed by a conclusion in section 6.

2.0 Importance of Height and Agility in Basketball

Figure 1 looks at the top five front court (Power Forward and Center) drafted each year. As can be seen the average height of the top drafter centers and power forwards since 1985. In 1985 the average height was just around 83 inches which is 6 ft 9 in. This dipped and then rose to its highest in 2015 at around 84 in and has since dipped to the lowest ever at 81 inches.

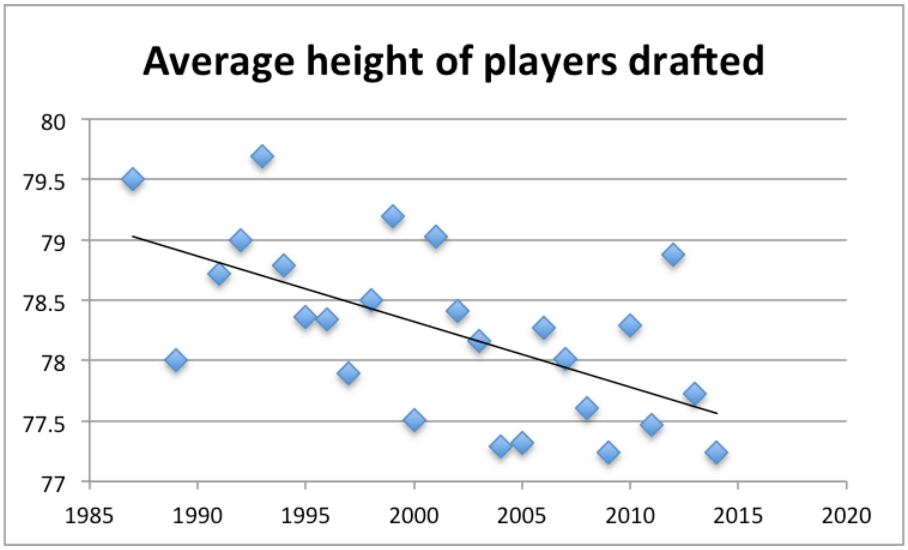
Figure 1: NBA Frontcourt Height



Source: FiveThirtyEight

Figure two looks at a similar trend line but looks at the entirety of all players regardless of position and finds that the same is true. The average NBA player is getting shorter. Part of this may be because there is a greater emphasis on big men like Dirk Nowitzki who although tall could shoot threes. There is less need for tall big men if they can not shoot threes and space the floor (not bunch up players next to the basket).

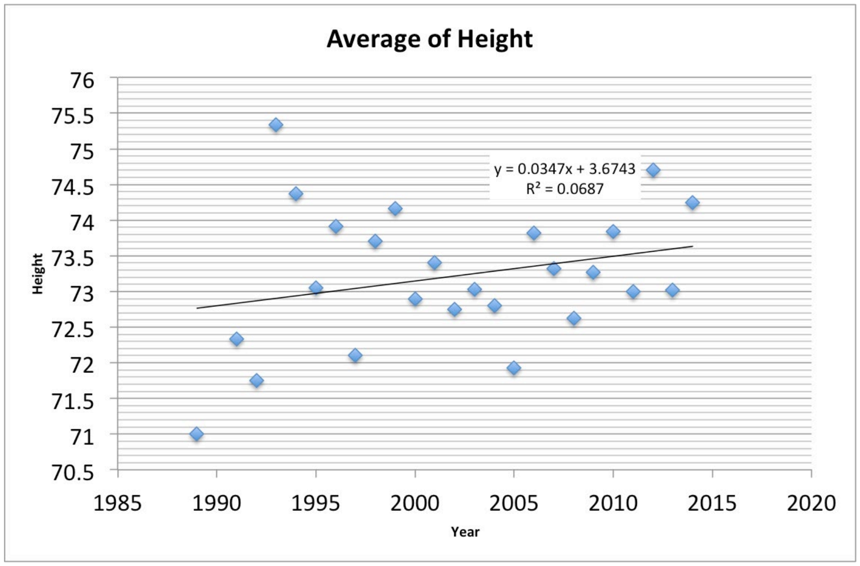
Figure 2: Average NBA height drafted



Source: UC Berkeley Sports Analytics

Figure three looks at height as well. UC Berkeley Sports Analytics found that at almost every position height is going down except one which was point guard. The average height has gone up from just around 72.5 in to now almost an inch taller at 73.5. The reason for this is most likely the same for centers. There is a need for taller players to have ball handling skills and if certain players are to short it will create mismatches on defense and on offense. One of the best examples of this is point guard Ben Simmons who is 6 ft 10 in but has incredible passing and ball handling skills.

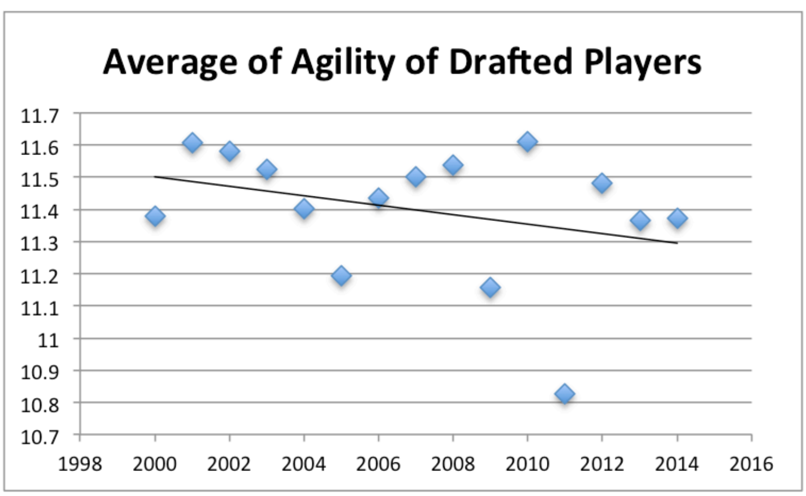
Figure 3: Point Guard Average Height



Source: UC Berkeley Sports Analytics

Figure four looks at the average agility which has gotten slightly faster over time. The agility tests have players start at the baseline, then run to the free throw line, defensively slide (shuffle essentially), then back pedal, then defensively slide again. This may be another reason why height overall has gone down because players have gotten faster, and the taller players are the slower they are and the more knew issues they have.

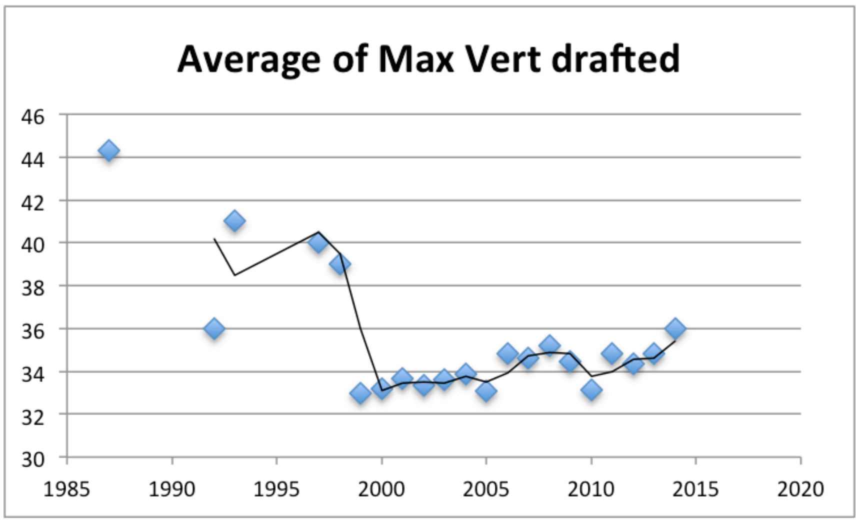
Figure 4: Average Agility of Drafted Players



Source: UC Berkeley Sports Analytics

Figure Five looks at the average max vertical over time. Originally it started out much higher but some of this is because the Draft Combine was not as popular. Over the last twenty years however it has gone up. This may also be why centers and power forwards have gotten shorter as those players may still be able to jump as high at those who are taller.

Figure 5: Average Max Vertical



Source: UC Berkeley Sports Analytics

3.0 LITERATURE REVIEW

The NBA Amateur Player Draft and drafts in general are pretty unique to North American sports and most of the research done on drafts and sports are relatively recent as there is a move towards using economics and modeling to gain an edge in sports. One of the earliest papers in 2011 aims to see if the draft is good at improving bad teams (Berri et al, 2011). One of the unique variables that the paper uses is the specific NCAA conference that the player competed in arranged as a dummy variable. There are several papers that look if the player played in a major conference, but not breaking them out separately. The found that where a player is drafted does not seem to tell much about performance in the NBA. In fact, they found the less then 5% of a players Wins Produces Per 48 Minutes (an all-encompassing stat) can be attributed to their draft position. Another paper developed in 2011 aimed at explaining a team's performance with revenues and aimed to fine the marginal revenue product of an individual player (Li, 2011). One of papers that inspired this paper looks at the success of players in which the author creates multiple models, however they use NBA minutes and NBA win shares during their third season. This paper did, however, shows many variables that will be collected for this paper as well as showing variables overlooked by current NBA GM's like turnover rate (Evans, 2018). The goal of this paper is to use some machine learning techniques as well and there were only a handful that looked at predicting NBA success. Predicting NBA Success: a Machine Learning Approach used a few really interesting techniques. The three techniques were a traditional Logistic Regression, a Support Vector Machine model, and a Random Forest Model (Kannan, 2018). The logistic regression model performed the worst but one thing that this paper will try differently is to also create a lasso logit, a lasso function will pick the variables in the model and therefore can create a much better model than a classic logit. The Random Forest model performed the best so this paper will also attempt to use the Random Forest method for several type of models. Another paper in 2015 aimed to do a similar paper however they used a much larger dataset from 1985 to 2005. One issue that this may cause is that there were more rounds in prior drafts as well as expansion teams that have been added which also affect the size of the draft (Greene, 2015). However, this seems not to be true when looking at the success of those players long term.

There are also several papers that look at other sports and leagues. One of the most prominent is Harris and Berri, Predicting the WNBA Draft: What Matters Most from College Performance. Similar to papers done on the NBA draft, there seems to a be strong significance with conference when it comes to where they will be drafted (Harris and Berri, 2015). The WNBA draft is a little different just because there are fewer teams so in theory the talent pool is even more selective. There are several papers that look at the NFL Draft. One recent paper looks at the difference between attributes correlated with higher draft positions and attributes correlated with NBA success. They find that many lower drafted picks outperform players at the same position drafted higher. The paper concludes the certain attributes are being overvalued in the draft and that also NFL teams are willing to take more risks in lower rounds. Another paper by Pitts and Evans looks at cognitive ability for quarterback drafted. They found a strong effect between NFL success and the Wonderlic Test which is like an IQ exam and is timed. Now the results of this paper actually contradict several others which found no correlation between NFL success and the Wonderlic Test, but this is the first paper to only look at Quarterbacks where often it is considered important for the player to be "smart" is a traditional sense because they need to memorize plays and have the ability to find the open man (Pitts and Evans, 2018). All of these papers have led to changes in how this paper will collect data as well as the models used and how the models will be implemented.

4.0 DATA AND EMPIRICAL METHODOLOGY

4.1 Data

The study uses panel data from 2006 to 2015. The idea was to give players three years to become a regular player within the NBA. Data were obtained from Basketball reference for NBA stats and college stats as well as data from NBA for draft combine stats. The year of 2006 was picked because high school draft recruits were able to skip college before 2006 and go right into the NBA. Foreign players who did not play in college (or U.S. players who played overseas) were ignored because of discrepancies in data between leagues as well as the differences in skill level between leagues. College level stats are the most recent year that person played instead of an aggregate because many players only play one year, as well as the fact the many senior who get drafted did not get big playing time till later in their collegiate careers. the Summary statistics for the data are provided in Table 1.

Table 1 Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
nbaG	264	348.8631	256.8395	1	1002
WingSpan	264	82.4786	3.830896	70.75	91
Height	264	78.9678	3.205999	70.25	85.75
MaxVerticle	264	35.27462	3.441991	25	44
Games Started	264	31.87692	7.187522	0	41
Effective Field Goal	264	0.541612	0.0482782	0.413	0.69
Player Efficiency Rating	264	23.23938	4.555071	11.65	37.15
College Points Per Game	264	15.77519	4.282486	3.39	28.65

4.2 Empirical Model

Using techniques that Harris and Berri, 2015 used in a similar model however it is changed with certain dummy variables for instance the use of big conference dummy.
The model could be written as follow:

$$\text{NBASuccess}=\beta_0+\beta_1\text{Pick}+\beta_2\text{BigFiveDummy}+\beta_3\text{Freshman}+\beta_4\text{Sophomore}+\beta_5\text{Junior}+\beta_6\text{GamesStarted}+\beta_7\text{FieldGoal}+\beta_8\text{Threesmade}+\beta_9\text{Freethrows}+\beta_{10}\text{Rebounds}+\beta_{11}\text{Assists}+\beta_{12}\text{Steals}+\beta_{13}\text{Blocks}+\beta_{14}\text{Turnovers}+\beta_{15}\text{PTSPERGAME}+\beta_{16}\text{DBLDBL}+\beta_{17}\text{TPLDBLE}+\beta_{18}\text{Wins(teamwins)}+b_{19}\text{WinShare}+\beta_{20}\text{EfectiveFG}+\beta_{21}\text{PlayerEfficencyRating}+\beta_{22}\text{Wingspan}+\beta_{23}\text{Height}+\beta_{24}\text{MaxVerticle} \quad (1)$$

NBASuccess is a binary variable that is derived from looking at if a player plater ¾ of their first three seasons and lasted in the league longer than three seasons. There were different papers that sued different metrics of success. The original plan for the variable was to use all-star picks but due to the limited number is created to small of a sample size. Several papers used games played and by turning it into a binary variable a logit model could be used

Independent variables consist of 24 variables obtained from college statistics and NBA draft combine. Appendix A and B provide data source, acronyms, descriptions, expected signs, and justifications for using the variables. First is pick which is what pick they are picked at between 1 and 60 per each draft. Second is a big five dummy which looks at whether the player played in a big five conference. The reason is to differentiate players who may have had similar stats but harder competition in one versus the other. The third, fourth, and fifth are dummy variables for grade when drafted. Sixth is games starts in college. Seventh, eighth, and ninth are field goal percentage, three-point percentage, and free throw percentage. Rebounds, Assist, Steals, and Turnovers look at those stats per a game. 15th is points per game in college. 16 and 17 are the number of double doubles and triple doubles over a season that they had. This is when a player has more than 10 units in two (or three for triple) respective categories in a game. This could be

10 steals and 10 points or 10 blocks or 10 rebounds in any combination. 18 is team wins that season. Win share is a stat that estimated number of wins a player produces throughout the season. Effective field goal is the same as field goal except it balances the fact that threes are worth more points. 21 is player efficiency rating which is a rating of a players per minute productivity. League average in per is always 15 so it is always balanced to that number. Lastly are physical stats, 22 is wingspan which is the measurement from fingertip to fingertip while arms are spread out. 23, is height with shoes on. Lastly is 24 which is max vertical which is how high someone can jump.

5.0 EMPIRICAL RESULTS

Three models were run in the regression results. The first is a normal regression that uses games played as the dependent variable. The next model is a logit model using the same variables. Lastly is a logit which is based on a lasso logit which is a machine learning technique that picks variables to enhance prediction accuracy. Below are the results. Due to the number of variables only variables that were either picked by the logit or significant are included. I was interested in the fact that although pick is significant in all the models, nothing else is significant across all level. The other interesting thing is that other then freshman, all the other variables are only significant at the .10 percent level. The R-Square remain constant around .3 one of the reasons that I think it is so low is because of the variables that are outside of the data given in a draft.

Table 2: Regression results

	Reg	Logit	Logit(based on Lasso)
Pk	-6.120***	-0.0800***	-0.0757***
Freshman	169.5**	0.769	
Junior	93.78*	0.348	
College Points Per Game	-6.133	-0.309*	
WinShare	15.86	0.46	0.235
Player Efficiency Rating	3.675	0.214	0.0518
WingSpanIn	-13.61	-0.246*	
Constant	460	10.27	0.335
R-Square	0.3011	0.318	0.2516

Note: *** , **, and * denotes significance at the 1%, 5%, and 10% respectively. Standard errors in parentheses

Interpreting these results shows that pick is the best indicator to see whether someone will have success within the NBA. Being a freshman seems to increase your chances of success but this is most likely due to that fact that players with the best talent put themselves in the draft as freshman and player who know they won't get drafted or have a low draft pick will wait until later year in college. I was surprised that many of the offensive stats were not significant in any of the models (which is why they are not included in the table above). Another interesting thing is what variables the lasso picked which include mostly calculated statistics.

In terms of the prediction results there are 4 different models. The first is a logit with only pick included to essentially give a bass line of how good NBA general managers are at drafting players. The next three are a logit, lasso logit, and random forest.

Table 3: Prediction results

	Just PK	Logit w/o Pick	LassoLogit	Random Forest
Sensitivity	84.39	86.05	71.61	100
Specificity	51.65	52.27	74.19	100

Correctly Classified	73.11	74.62	71.92	100

For all of these models, draft pick is not included to see if other stats could predict success. In addition, no validation data was set partially because the data was so small and would potentially skew the models. As can be seen the only outlier is the random forest which did fantastically. One of the reasons this might be is it picks through random noise very well. Another interesting fact is that other than the random forest the Logit does the best, but both just using the pick and the logit have real issues with specificity. The most balanced model is the lasso logit which has the worse correctly classified rate, but the best specificity mode.

The takeaways is the NBA general managers are very good at drafting talent and this can be seen in both the models which shows that picks is a highly significant variable in determining NBA success as well as in the predictions where it performs similarly to the other models. Players who are freshman also seem to do better as well. One of the more interesting things is that points per game is negative which points to an idea the players who score a bunch of points may not have high success. This may be because of play style in college or it may be because having to many points per game could point to selfishness or lacking attributes in other areas.

5.0 CONCLUSION

In summary, although these models were overall good, and the prediction results were expected but very good. However, there were some issues, the first is that the model disagreed on what variable were significant and mattered. In addition, the models had low R-Squares. One of the reasons this may be is because the data itself is lacking. Not everyone goes to the NBA draft combine, so many of the stats this paper originally planned on using become unusable and had to use more basic height stats. Another reason is that basketball IQ is often thrown around for players who make smart passes and such, but there is no measure for that in the NBA draft. It is unclear whether teams do that sort of analysis in meetings with players and in work outs. Overall, this shows that general managers are performing at what a model would expect them to perform at and shows no faults at least as whole. Surely there are some General Managers who perform better then others, but general managers perform very well at predicting talent that will succeed.

Appendix A: Variable Description and Data Source

Acronym	Description	Data Source	Expected Sign

PK	Pick in each respective draft, lower is better	Basketball Reference	-
BigFiveDummy	Coded if player played in a major five conference	Basketball Reference	+
Freshman	Dummy if drafted as a freshman	Basketball Reference	+
Sophomore	Dummy if drafted as a Sophomore	Basketball Reference	-
Junior	Dummy if drafted as a Junior	Basketball Reference	-
GS	Games Started previous season in college	Basketball Reference	+
FG	Field Goal Percentage in college	Basketball Reference	+
3P	Three-point percentage in college	Basketball Reference	+
FT	Free Throw Percentage in college	Basketball Reference	+
TRB	Rebounds per game in college	Basketball Reference	+
AST	Assists per game in college	Basketball Reference	+
STL	Steals per game in college	Basketball Reference	+
BLK	Blocks per game in college	Basketball Reference	+
PTSPerGame	Points Per Game in college	Basketball Reference	+
StDBLDBL	Number of Double Doubles (hitting 10 or more units in two categories; steals, blocks, points, rebounds, Assists)	Basketball Reference	+
StTPLDBL	Number of Triple Doubles (hitting 10 or more units in two categories; steals, blocks, points, rebounds, Assists)	Basketball Reference	+

Wins	College Team Wins	Basketball Reference	+
WS	Win share (calculated field that projects player worth)	Basketball Reference	+
PER	Player Effienceny Rating (Calculated field that displays per minute rating)	Basketball Reference	+
eFG	Adjusted field goal percentage to account that three pointers are worth more	Basketball Reference	+
WingSpan	Wingspan in inches	NBA.com	+
Height	Height with shoes on in inches	NBA.com	–
MaxVerticle	Jumping Ability	NBA.com	+

BIBLIOGRAPHY

Berri, D. J., Brook, S. L., & Fenn, A. J. (2011). From College to the Pros: Predicting the NBA Amateur Player Draft. *Journal of Productivity Analysis*, *35*(1), 25–35. https://doi-org.bryant.idm.oclc.org/https://link.springer.com/journal/volumesAndIssues/11123

Evans, B. A. (2017). From college to the NBA: what determines a player's success and what characteristics are NBA franchises overlooking? *Applied Economics Letters*, *25*(5), 300–304. doi: 10.1080/13504851.2017.1319551

Greene, A. C. (2015). The Success of NBA Draft Picks: Can College Careers Predict NBA Winners? *St. Cloud State University*. Retrieved from https://repository.stcloudstate.edu/cgi/viewcontent.cgi?article=1002&context=stat_etds

Harris, J., & Berri, D. J. (2015). Predicting the WNBA Draft: What Matters Most from College Performance? *International Journal of Sport Finance*, *10*(4), 299–309. https://doi-org.bryant.idm.oclc.org/http://fitpublishing.com/journals/ijsf

Kannan, S. (2019, June 30). Predicting NBA Rookie Stats with Machine Learning. Retrieved from https://towardsdatascience.com/predicting-nba-rookie-stats-with-machine-learning-28621e49b8a4

Li, H. (2011). An Analysis of On-Court Performance and Its Effects on Revenues. *University of California, Berkeley*. doi: https://www.econ.berkeley.edu/sites/default/files/li_harrison.pdf

Pitts, J. D., & Evans, B. (2019). Drafting for Success: How Good Are NFL Teams at Identifying Future Productivity at Offensive-Skill Positions in the Draft? *American Economist*, *64*(1), 102–122. https://doi-org.bryant.idm.oclc.org/http://aex.sagepub.com/content/by/year