Difference by Conference in NBA Draft Status Among Power Five Collegiate Basketball Players from 2011-2021

Abstract

This paper investigates the differences in NBA draft status among Division I college basketball players who play in the group of conferences known as the “Power Five.” Using a comprehensive data set with simple and complex statistics for every player who played Division 1 collegiate basketball from 2011-2021, I examine factors such as player performance metrics, demographic information, and conference strength to identify the determinants of NBA draft status. I specifically look at the differences in producing NBA draft picks between the five conferences of interest. The findings of this analysis show a small difference in the ability to produce NBA draft picks between the Power Five conferences and lead to some interesting questions that warrant further analysis. This analysis offers some insight into the interplay between collegiate and professional sports and contributes to the discussion on the complex interaction between where an athlete chooses to play collegiate sports and their associated professional outcomes.

Introduction

The National Basketball Association (NBA) is a multi-billion dollar organization that is internationally known and employs what are widely considered the best basketball players in the world (Forbes, 2023). A spot on an NBA roster carries a minimum salary of $1.1 million (Statista, 2024). This significant sum for simply making the NBA is a strong source of motivation for numerous basketball players. Since each NBA player is worth at least $1 million, scouting and drafting the right player is crucial. There is no clear consensus on what makes a top NBA prospect. Some studies emphasize traditional metrics such as scoring and rebounding while others focus more heavily on what is colloquially referred to as the *eye test* (Sailofsky, 2018). The *eye test* refers to the more subjective opinion scouts form by watching a player on and off the court. Modern scouting philosophies tend to apply a blend of these two viewpoints; on-court performance matters but should be considered in the context of how a player interacts with those around him and where he plays (Larkin et al., 2022).

One key aspect of player development that is missing in the current literature is the impact college conferences have on NBA draft status. This analysis attempts to fill in this gap and analyze the impact certain Division 1 conferences have on helping players get drafted. This knowledge is important as it provides empirical evidence for a question both fans and analysts often debate: Which conference is the best at developing talent? It also provides evidence that may be useful for a prospective player looking to optimize their chances of getting drafted into the NBA.

Conceptual Model

There is more to an NBA prospect than just their on-court performance. While individual statistics matter, the level of competition a player plays against provides important context that can raise a player's profile. Different NCAA conferences receive different levels of funding and play different levels of competition. The Power Five conferences are widely considered to be the most competitive conferences and thus receive a large influx of revenue every season from television deals, sponsorship deals, and booster club donations. Most data used to justify conference supremacy focus solely on team success, not player success. This analysis seeks to examine the effect each Power 5 conference has on the probability of achieving the ultimate level of individual basketball success: getting drafted into the NBA.

Data

The data used for this analysis was collected from official NCAA data sources by Aditya Kumar and accessed on the data repository website *Kaggle*. The original dataset features season-wide data for all Division I collegiate basketball players from 2009-2021 and has 25,719 observations. Every season a player plays in college is given its own observation which results in a player having between one and four corresponding observations. This data consists of demographic information such as the height of the player and the school they attended, as well as both simple and advanced on-court statistics. These include simple statistics such as points scored per game and rebounds per game, as well as advanced metrics such as effective field goal percentage.

The sample used in this analysis consists solely of players who played in the interval 2011-2021 and who played in either the Atlantic Coast Conference (ACC), Big Ten Conference (B10), Big Twelve Conference (B12), Pacific 12 Conference (PAC12), or Southeastern Conference (SEC). To correct for multiple observations of the same player, the observation corresponding to each player’s final season was kept while any other seasons were dropped. This was done because of the belief in recency bias; a player's last year before declaring for the NBA draft is often the year viewed by scouts as the most important. After creating this sample, 21,970 observations were dropped resulting in a sample size of 3,749 with 30 variables.

The variables used in the sample are a mix of demographic, simple, and advanced variables. The primary dependent variable is the *pick* variable which is a binary indicator of whether or not a player was drafted into the NBA. The primary independent variable is the *conference* variable which is a categorical variable with five levels that correspond to each of the Power Five conferences. The ACC is used as the reference category for this analysis. The other variables in this sample are measures of simple and advanced statistics. The simple statistics are common basketball measures such as *points per game*, *assists per game*, and *defensive rebounds per game*. The advanced statistics are measures such as *effective* *field goal percentage*, and *minutes played*. A complete list of all variables and their corresponding summary statistics can be found in Table 1.

Table 1: Summary Statistics

	Mean	Std. Dev.	Min	Max
pick	**0.102**	**0.304**	**0**	**1**
Min_per	38.42	30.96	0.1	96.7
usg	18.44	6.93	0	49.2
eFG	44.72	21.35	0	150
TS_per	47.81	20.77	0	150
ORB_per	5.84	7.14	0	122.8
DRB_per	13.266	9.057	0	115.1
AST_per	11.244	10.503	0	100
TO_per	18.58	13.89	0	100
FT_per	0.56	0.31	0	1
twoP_per	0.42	0.22	0	1
TP_per	0.24	0.20	0	1
blk_per	2.08	3.76	0	96.4
stl_per	1.75	2.52	0	61.6
year	2017	3.32	2011	2021
GP	22.73	11.60	1	40
stops	91.36	78.12	0	360.57
mp	17.31	11.99	0.33	39.32
oreb	0.81	0.83	0	4.19
dreb	2.08	1.75	0	9.71
ast	1.19	1.36	0	9.75
stl	0.55	0.51	0	3.44
blk	0.35	0.52	0	4.65
pts	6.47	5.81	0	27.38
ht	77.52	3.58	67	8

Conference	# of Players	Frequency
ACC	843	22.58
B10	761	20.38
B12	600	16.07
PAC12	695	18.61
SEC	835	22.36
Observations	3,734	

Methods

To examine the difference in draft status among Power Five collegiate basketball players by conference, a probit model was used. A probit model was used to accommodate the binary response variable *pick* and to produce the two main values of interest: the marginal effects of each conference and the predicted probability of an average player in each conference being drafted. The main question of interest was examining the difference in draft status by conference. To answer this question, marginal effects were analyzed to determine a conference's impact on the predicted probability of a player getting drafted if they played in a different conference other than the ACC. After examining the marginal effects, the predicted probability of getting drafted for a player with modal values in each conference was calculated to provide an additional measure of how good each conference is at developing NBA-level talent.

To ensure the best model was used, four models with different functional forms were examined. The first functional form was a base model where no interactions or transformations were conducted. The second functional form includes an interaction between the *points* and *assists* variables to account for the joint nature of the two statistics. The third functional form features the interaction between *points* and *assists* as well as an interaction between the *defensive rebound* and *offensive rebound* variables. The last functional form features the same interactions as the third functional form and includes an additional interaction between *blocks* and *steals*. Each of these interactions are between similarly related statistics and is included to account for the related nature of the statistics and their respective impacts on whether or not a player gets drafted. The model used for the analysis was chosen based on the Akaike information criterion (AIC) goodness of fit measure. Model four had the lowest AIC and thus was chosen as the best model (see Table 2). Lastly, a Wald test was conducted to examine the joint significance of the interaction effects included in model four. The resulting test was statistically significant at the $p < 0.05$ level which indicates that there is evidence to conclude that the interaction terms collectively contribute significantly to explaining the dependent variable in the model. Because of the results of the goodness of fit measure and the specification test, model four was used for analysis.

Table 2: Information Criteria Results (Preferred Model in Bold)

Model	AIC
No Interactions	1428.67
Interaction between *points* and *assists*	1430.48
Interaction between *points* and *assists* and interaction between *defensive rebounds* and *offensive rebounds*	1420.08
Interactions between *points* and *assists*, *defensive rebounds* and *offensive rebounds*, and *blocks* and *steals*	**1414.29**

One major concern during this analysis was the problem of endogeneity. One variable that appeared to be endogenous was the *minutes played* variable. The number of minutes a player plays is determined by a wide variety of factors that are both quantitative and qualitative. Quantitatively, players with better statistics should play more although that is not always true. Qualitatively, coaches may allot minutes based on unknown factors such as personal feelings or situational effects. Many of these factors were not included in the model and the *minutes played* variable appeared to be correlated with the error term. To address the problem of endogeneity, the *eprobit* command in Stata was used with the demographic variable *height* used as an instrumental variable. This variable was chosen as an instrument because the height of a player is generally correlated with the minutes a player plays due to the increased physical size needed to play certain positions in basketball. Furthermore, a player's height does not have a direct effect on whether a player gets drafted in and of itself, only through its correlation with minutes played and the associated statistics that come from playing more or less minutes in a game. The *eprobit* command has a built-in test for endogeneity which was statistically significant which implies that the *minutes played* variable is endogenous and that the *eprobit* model is the correct model to use.

Results

After running the probit model and accounting for endogeneity, the following results were found. The marginal effects for each of the conferences were negative and small with only the result for the Big Ten conference being statistically significant (see Table 3). The marginal effects for every conference except for the SEC were negative. This implies that on average, playing in the Big Ten, Big 12, or Pac 12 instead of the ACC is associated with a slight percentage point decrease in a player's predicted probability of getting drafted, while playing in the SEC instead of the ACC is associated with a slight percentage point increase in getting drafted.

Table 3: Probit Model Output

	Coefficients	AME
ACC	0	0
	(.)	(.)
B10	-0.039	-0.027*
	(0.021)	(0.012)
B12	-0.014	-0.010
	(0.018)	(0.013)
PAC12	-0.016	-0.012
	(0.017)	(0.012)
SEC	0.0028	0.0022
	(0.016)	(0.012)
pts	0.026**	0.017***
	(0.0084)	(0.0014)
ast	0.0452*	0.0247***
	(0.0200)	(0.0042)
c.pts#c.ast	-0.00078	-
	(0.0011)	
blk	-0.0056	0.0411***
	(0.0192)	(0.0091)
stl	0.051*	0.068***
	(0.025)	(0.011)

	Coefficients	AME
c.blk#c.stl	0.064*	-
	(0.029)	
dreb	0.0252	0.0085
	(0.0143)	(0.0045)
oreb	0.0318	-0.0024
	(0.0252)	(0.0072)
c.dreb#c.oreb	-0.0089	-
	(0.0053)	
efg	0.0029*	0.0021***
	(0.00114)	(0.00051)
mp	0.078***	0.055**
	(0.0031)	(0.0183)
ht		0
		(.)
_cons	-2.133***	
	(0.21)	
N	3734	3734

Standard errors are in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The second result found was the predicted probability of a player with modal values getting drafted in each conference. The predicted probabilities were calculated using the *predict* command in Stata. The ACC had the highest predicted probability at 12.11% while the Big Ten had the lowest predicted probability at 8.21% (see Table 4). The mean squared error (Brier score) was calculated to provide a measure of precision for these estimates. Brier scores range from zero to one and a lower Brier score implies a more accurate prediction. The Brier score value calculated was 0.058 which indicates an accurate predictive performance.

Based on the analysis conducted, there is a small difference in NBA draft status between Power Five conferences. The ACC has the highest predicted probability of being drafted for a player with modal values and the marginal effects for every conference other than the SEC are negative which indicates that choosing to play in either the Big Ten, Big 12, or Pac 12 conferences instead of the ACC decreases a player’s probability of getting drafted to the NBA.

Table 4: Predicted Probability of Getting Drafted for an Average Player by Conference

Conference	Predicted Probability
ACC	12.11%
B10	8.21%
B12	9.85%
PAC12	10.08%
SEC	11.02%

Discussion

The main goal of this analysis was to examine the difference in how well each Power Five conference does in producing NBA draft picks. Each of the Power Five conferences had predicted probabilities very similar to each other with small marginal effects. These measures show the competitiveness and shared prestige the Power Five conferences have; choosing to play in any of the Power Five conferences does not impact the probability of getting drafted by much. Further analysis comparing these results to results from an analysis of non-Power Five conferences would provide a more robust look at which conferences are better than others at producing NBA talent.

Although the findings of this analysis are small in scale and in many cases not statistically significant, the findings are still useful in advancing the discussion regarding the intersection between player choices and future career outcomes. It is rational to assume that top-tier high school basketball players desire to choose a college program that would maximize their chances of playing in the NBA. While numerous factors such as geography, family history, and playing style affect which college a prospective player chooses, a major driving force is the prestige of the school. A major factor that adds prestige to a collegiate basketball program is the conference they play in. This analysis provides a quantitative result that provides some evidence to differentiate the effectiveness of each Power Five conference in producing NBA draft picks. This information would be valuable to both prospective collegiate players seeking to maximize their career potential as well as collegiate coaching staff seeking to attract top talent.

There are numerous limitations to this analysis. The first main limitation comes from the analysis itself. The problem of endogeneity was noted and addressed but endogeneity may play a larger role in basketball data than captured in this analysis. Basketball statistics are highly correlated in nature due to the rules of the game. This analysis may not have captured all endogeneity between variables in the sample and further work may be done to ensure that the estimates calculated are not biased in ways unperceived by the author. Similarly, the results found by examining only Power Five conferences may not be generalizable to the entire population of Division 1 basketball players. There are 32 Division 1 conferences, each with various levels of prestige, revenue, and exposure that can impact the chances a player has of getting drafted. This analysis would benefit from further study comparing all conferences and their respective marginal effects and predicted probabilities. Lastly, this analysis examines a period before NCAA regulations regarding player transfers were updated. Further studies on this subject should take into account the ever-changing landscape of NCAA rules and regulations to ensure accurate and usable results.

Overall, there is little difference between the Power Five conferences in producing NBA draft picks. An average player in the ACC has the highest predicted probability of getting drafted out of the five conferences and choosing to play in either the Big Ten, Big 12, or Pac 12 instead of the ACC is associated with a slight percentage point decrease in the probability of getting drafted. However, choosing to play in the SEC instead of the ACC results in a slight percentage point increase in the probability of getting drafted. Every player and situation is unique, but based on the results of this analysis alone, a prospective collegiate basketball player seeking to maximize their chances of making it to the NBA may want to consider playing in either the ACC or SEC.

Appendix

Stata Code: set more off

```
clear
capture log close

log using "Paper.log", replace

* Data Cleaning

encode conf, gen(conf_numeric)
drop conf v18 v32 v33 v34 v35 v36 v37 v38 v39 v40 v41
drop if ht <=59
drop player_name
drop team
drop num

* Create Summary Stats Table and Find Modal Values

estpost summarize *

esttab using "Paper.rtf", cells("mean sd min max") ///
      label nonumber collabels("Mean" "Std. Dev." "Min" "Max") ///
        replace



* Run Probit Model and assess different functional forms

* Form 1

probit pick ib(1).conf_numeric pts blk stl ast dreb oreb efg mp
estat ic


* Form 2

probit pick ib(1).conf_numeric c.pts##c.ast blk stl  dreb oreb efg mp
estat ic

* Form 3

probit pick ib(1).conf_numeric c.pts##c.ast blk stl  c.dreb##c.oreb efg mp
estat ic

* Form 4

probit pick ib(1).conf_numeric c.pts##c.ast c.blk##c.stl  c.dreb##c.oreb efg mp
estat ic
```

```
* Wald Test

test c.pts#c.ast c.blk#c.stl  c.dreb#c.oreb


* Now that I have identified the functional form with the lowest AIC:

* To deal with endogeneity of mp variable, use height as an IV, then run eprobit

eprobit pick ib(1).conf_numeric c.pts##c.ast c.blk##c.stl c.dreb##c.oreb efg, endog(mp=ht)
estimates store reg1


margins, dydx(*) post

estimates store reg2

esttab reg1 reg2 using "Paper_Margins.rtf", replace


predict phat, pr

summarize phat
sort conf_numeric
by conf_numeric: summarize phat, meanonly

egen mean_phat = mean(phat), by(conf_numeric)

list mean_phat

gen brier_score = (pick - phat)^2

* Calculate mean of squared differences (Brier score)
sum brier_score


log close
```

References

Five ACC teams earn bids to NCAA tournament. Five ACC Teams Earn Bids to NCAA Tournament - Atlantic Coast Conference. (n.d.). https://theacc.com/news/2024/3/17/mens-basketball-five-acc-teams-earn-bids-to-ncaa-tournament.aspx#:~:text=Current%20ACC%20schools%20have%20won,67%20Men's%20Final%20Four%20appearances.

Kumar, A. (2021, November 4). *College basketball 2009-2021 + NBA Advanced Stats*. Kaggle. https://www.kaggle.com/datasets/adityak2003/college-basketball-players-20092021/data

Larkin, P., Sanford, M., Talpey, S., Gorman, A. D., & Reeves, M. J. (2022, December 28). *Talent identification in youth basketball: Talent scouts' perceptions of the key attributes for athlete development*. Human Kinetics. https://journals.humankinetics.com/view/journals/iscj/10/2/article-p163.xml

Minimum NBA salary 2024. Statista. (2024, February 13). https://www.statista.com/statistics/1009569/minimum-nba-salary/#:~:text=The%20minimum%20salary%20for%20players,over%20925%20thousand%20U.S.%20dollars.

Sailofsky, D. (n.d.). *Drafting errors and decision making theory in the NBA draft*. Drafting Errors and Decision Making Theory in the NBA Draft. https://dr.library.brocku.ca/bitstream/handle/10464/13452/Brock_Sailofsky_Daniel_2018.pdf?sequence=1&isAllowed=y

Teitelbaum, J., & Ozanian, M. (n.d.). *NBA valuations 2023*. Forbes. https://www.forbes.com/lists/nba-valuations/?sh=5e16b5176982

Wilco, D. (2018, October 5). *The best conferences in NCAA tournament play since 1985*. NCAA.com. https://www.ncaa.com/news/basketball-men/article/2018-10-05/best-conferences-ncaa-tournament-play-1985