A Repeated Cross-Sectional Analysis of College Basketball Player Outcomes and the Impact of an Additional Year of Play at the College Level on NBA Draft Position and Career Earnings

David Selemba and Kira Hegdal

November 2023

Abstract

This research focuses on an empirical regression approach to determine whether number of years playing college basketball can be an explanatory variable to response variables NBA draft pick number and career earnings. Performing regressions between players entering the draft after their sophomore year versus freshman year, and entering after their senior year versus their junior year, our study revealed a negative correlation between the number of years spent playing NCAA basketball and NBA draft selection. That is, staying in college longer results in being picked later in the draft. We estimate that younger players with more time to develop under NBA coaches seem to be more attractive to teams when selecting their draft picks than players who have developed their skills more at the collegiate level. Regarding lifetime earnings, we found that sophomores entering the draft earn less than their freshman counterparts, but seniors and juniors earn more than those freshmen, on average, but this was not statistically significant. This may be due to the impact of age when entering the NBA or other factors we could not account for. There may be other factors playing into the decisions of teams and coaches when looking at the selection pool for the draft based on private information or selection bias, but our analysis did find statistical significance with players who have played less time in college being chosen over players who played more years in college.

Introduction/Background

Exploring college basketball outcomes via empirical economic analysis is a modern avenue of research with significant implications and relevance to college students, potential players, and sports fans. The main question driving our research is whether an additional year of college basketball increases the likelihood of being drafted into the National Basketball Association (NBA) and impacts the subsequent career earnings. Understanding the factors that influence the draft selection and career trajectories of college basketball players can demonstrate the value of collegiate experience in shaping success in the professional scene. The decision-making process of players on the cusp of being drafted, whether to return to college for another year of play or enter the NBA Draft, reflects the combination of individual skill development, career strategy, and the perceived advantages of additional collegiate experience. This study aims to examine these dynamics and contribute valuable insights to the fields of sports economics and labor market outcomes.

The NBA Draft is an annual event for NBA teams to select new players for their rosters. The draft consists of two rounds, with 30 players chosen each round (one pick for every NBA team). The draft order is determined through a lottery for non-playoff teams, and based on regular-season records for playoff teams. The selection process involves evaluations of players' skills, potential, and fit within a team's short-term and long term strategy to reach their goals. Teams review college performance, conduct private workouts, and consider mock drafts to make these decisions.

If an additional year of college basketball is found to have a significant impact on NBA prospects and earnings, it could inform discussions on the value of extended education or training periods in other professional domains. This research may also provide insights into how institutions can better support athletes in navigating critical career decisions.

This research contributes by leveraging the toRvik database, which offers a comprehensive set of player statistics spanning back to 2010. This dataset allows for a detailed analysis of the career paths of players over multiple NCAA seasons. The inclusion of factors such as starting salary, mock draft ranking status, and draft order provides a complex understanding of success metrics beyond mere draft selection.

Literature Review/Previous Research

Several papers have delved into related topics, examining the relationship between college sports participation and professional success. For instance, Groothius, et al. (2007) investigated early entry into the NBA draft and its implications on player selection and career trajectories. They utilized two competing models to draw their conclusions—the human capital model and the Lazear (1995) option value model—to determine why teams opt for early selection under the revised rookie contract system. Using a comprehensive panel study covering NBA players from 1989 to 2002, they looked at individual player performance, salary data, and draft positions over a 12-year period. They concluded that teams may select players early to capitalize on potential upside, aligning with the option value hypothesis, and that players entering the NBA early exhibited accelerated skill development, playing fewer minutes in their first year compared to players who played all four of their years in college. The researchers suggested a willingness among teams to select players with potential for growth over immediate skill proficiency.

Greene (2015) investigated the predictive factors influencing success in the NBA, focusing on players from the 1985 to 2005 drafts who had at least one year of NBA experience and played in Division I NCAA. He analyzed both first and second-round picks, totaling 841 players. The key performance metrics measured included Player Efficiency Rating (PER), win shares, and win shares per 48 minutes. Taking into account draft pick position, college statistics, and physical attributes like height and weight, Greene created a model to predict player success. Ultimately, his statistical analysis of the college statistics effectively predicted performance, particularly in terms of win shares per 48 minutes. However, this predictive power did not significantly enhance performance predictions for first-round draft picks when considering PER and win shares.

Another senior thesis, Wolfe (2022), analyzes both the effectiveness of NBA organizations in player drafting and whether college experience significantly influences a player's early NBA career performance in their first three seasons. Data was gathered by measuring the first 3 years of a player’s performance in the NBA through the usage of the advanced statistic Win Shares. His research led him to conclude that college experience does not play a substantial role in shaping a player's early career performance in the NBA and suggested that other factors may be influencing the drafting process in the NBA.

Theoretical Issues/Hypotheses

The dataset and research question we have chosen raises interesting theoretical issues, particularly in the evaluation of top draft picks such as Deandre Ayton, James Wiseman, and Tyrese Haliburton. Given their limited college playing time, there's a potential for inflated or deflated college statistics, making it challenging to gauge their true performance capabilities. This raises questions about the reliability of traditional college metrics in predicting NBA success for players with shorter college careers. These outliers highlight the importance of alternative evaluation methods, such as high school tapes, in assessing the overall skills and potential of these players. These are not included in the data we selected, so we may be limited in drawing conclusions on such players. Additionally, the mention of undisclosed statistics that are not available to the public adds some more complexity, pointing to potential unexplored variables that could significantly impact player evaluation. This introduces a selection bias into the analysis, as players choose whether to enter the draft or not, even if it may not be the best long-term option. In addition, it also reflects the differences between mock draft analysts evaluations, which have been used in this analysis, and team intel that likely hinders their accuracy and limits this data as an estimation tool. Navigating these theoretical challenges is crucial for refining hypotheses and ensuring a more comprehensive understanding of the factors influencing NBA success beyond conventional college statistics.

The hypotheses we will be testing in the context of what we have learned in class this year are as follows. The null hypothesis to gauge draft positions is that there is no significant difference in NBA draft positions between players who choose to play an additional year in college and those who enter the draft early. The alternative hypothesis is players who spend an additional year in college will, on average, have a higher NBA draft position compared to those who enter the draft early. When looking at starting salaries, the null hypothesis we will test is that the starting salaries of NBA players are not influenced by an additional year of college basketball experience. The alternative to this hypothesis is that players who spend an extra year in college will, on average, earn higher starting salaries in the NBA compared to those who enter the draft early.

Description of the Data with Summary Statistics

To collect our data, we used a database from the R package toRvik, developed by Bart Torvik to provide free and accessible data about college basketball players including over 40 variables for each player, the team statistics, and coaching statistics. We only looked at individual player statistics for our model, but the database has a lot of resources that could be used in further research. The data we are using is repeated cross-sectional, as player statistics are recorded each year continually that they play in the NCAA. There are countless statistics available to us within this wide data, but we focused mainly on counting player stats as well as efficiency and value evaluations to give control for player performance. By using counting stats to show how the players actually performed when on the court as well as the advanced numbers to adjust between differences in playing time and schedules, we will be able to account for nearly all of their college production in our models. We also used data from two separate Github repository projects by NBA analysts. One contained salaries earned by every player in the NBA from 2000-2022, which we in turn summed to find a player's lifetime earnings while in the NBA. The other compiled a consensus of upwards of 20 mock drafts as well as NBA combine performance to create a statistical grade for a player's draft value. A higher grade would indicate that a player has a higher probability to succeed based on expert consensus and would in turn have a higher incentive to declare for the draft. Many mock drafts are paywalled or archived from drafts dating back as far as 2011, so this model was the best available option but could be improved upon with further resources and time. The descriptors for all of the variables in the models are listed in the following table:

Data Descriptor Table

Variable	**Description**
porpag	Points over Replacement per adjusted game
dporpag	Defensive PORPAG
adj_oe	Matchup adjusted offensive efficiency
ajd_de	Matchup adjusted defensive efficiency
bpm	Box plus-minus
ppg	Points per game
rpg	Rebounds per game
apg	Assists per game
exp	Experience level in college dummy variable; = 1 if player is Sophomore/Senior
high_grade	Mock draft consensus score; = 1 if player is above cutoff for high score
pick	NBA Draft selection number
lifetimeearnings	Sum of career earnings

In addition to the descriptors, a statistical glimpse of the toRvik box score and efficiency stats have been included below.

Statistic	N	Mean	St. Dev.	Min	Max
porpag	1,294	3.30	1.56	-1.89	7.50
dporpag	1,294	3.36	0.92	0.53	5.62
adj_oe	1,294	118.30	12.35	52.83	152.85
adj_de	1,294	92.78	6.41	73.21	115.74
bpm	1,294	5.84	3.33	-9.08	18.67
ppg	1,294	12.61	4.97	1.12	29.31
rpg	1,294	5.33	2.35	0.41	14.38
apg	1,294	2.14	1.54	0.07	9.94

As we can see, with twelve years of data we have upwards of 1000 observations in the data set just for players who have been drafted into the NBA during this period. It is interesting to see the much larger standard deviation value on adjusted offensive efficiency in comparison to defensive efficiency, although this is likely in part due to strong outlying values for superstar and defensive specialists as well as the weights in the stat. In addition, it also stands out that some drafted players even had seasons in college in which they actually performed below replacement level, evidenced by negative box plus-minus and PORPAG. This is of note as we will look to find if these players spent extra years in college developing their game or if they were drafted off other factors not identified in this study.

Empirical Model/Estimation Method

We split the data by the number of years played in college and decided to focus on comparing only players who were drafted into the NBA with a declaration season that is one year of experience apart. More specifically, we will study the difference between declaring as a freshman and as a sophomore, as well as juniors and seniors. The data was also grouped by position to identify any trends that may be important, although toRvik had a wide array of player positions that functioned more as their role in college. This made it more challenging to use statistically and was ultimately excluded from the models due to its arbitrary nature.

Additionally, we decided to exclude the top 14 players, considering the impact of the lottery and the presence of outliers, our focus shifts to the distinctive group comprising the lower half of the first-round picks and the upper echelon of the second-round picks. This strategic grouping aims to pinpoint the territory between the top-tier selections and those who narrowly missed the cut. Additionally, we have chosen to eliminate undrafted players, as they are irrelevant to our specific research question. By honing in on the intersection of first/second-round considerations and the borderline scenario of not being drafted at all, we will get a more targeted analysis of player outcomes on the margins within our research framework.

To compare players who entered the draft at different times in their career, we are using an Ordinary Least Squares (OLS) regression estimation model to compare statistics of each group's final year in college. If a player enters the draft as a sophomore, they will be considered the treatment variable and those that declared as freshman are used as the control dummy. We compared all experience levels to determine the effect of each additional year of play at the college level on the draft position and earnings. We will look at offensive and defensive efficiency as well as total box score statistics to evaluate a player's college statistics. In addition, we also use a regression discontinuity analysis using the mock draft grades to identify if an interaction between a stronger grade and declaring at an older age has a significant effect on players draft selection as well as their second contract earnings. We decided to run the earnings discontinuity on second player contracts due to the fact that these are often the most impactful financially for players, as all rookies picked in the late first and early second rounds earn a similar salary during their first few years in the league. This provides a more long-term analysis of the effect of staying in college for an additional season, as these players may see more expected value from their choice to stay in school later on in their careers.

We chose to use an OLS Regression for a few reasons–it is widely used in statistical analysis so it is accessible and interpretable to a large audience, and it is the most appropriate regression model to use when examining linear relationships between variables, which we wanted to analyze with our explanatory variable of years played of college basketball and response variables draft pick and career earnings. We also added a regression discontinuity analysis of pick onto mock draft rankings, which allows us to identify whether there is a significant and abrupt change in draft pick around a specific threshold, like being a 3-star versus 4-star rated recruit, and allow us to draw conclusions about how player evaluations may influence the actual draft selection.

Discussion of Results

When running the linear regression model of the players who declared for the draft after their freshman year vs one and done's, we first decided to run the model on pick in the draft. The formula for the regression was as follows.

$$pick = \beta_1 porpag + \beta_2 dporpag + \beta_3 adjoe + \beta_4 adjde + \beta_5 bpm + \beta_6 ppg + \beta_7 rpg + \beta_8 apg + \beta_9 sophomore + u$$

The same regression was run identically, only substituting seniors into the regression experience variable. This will be regression (2) in Table 1 below. The results were as follows:

Table 1: Pick OLS Dummy Regressions

	Dependent variable:	
	pick	
	(1)	(2)
porpag	-2.822	-4.862*
	(2.590)	(2.566)
dporpag	-2.352	0.515
	(2.327)	(2.595)
adj_oe	-0.064	0.223
	(0.243)	(0.284)
adj_de	0.173	0.591*
	(0.255)	(0.307)
ppg	0.026	0.071
	(0.336)	(0.322)
rpg	0.021	0.192
	(0.473)	(0.397)
apg	1.765***	0.173
	(0.659)	(0.627)
as.factor(exp)So	5.615***	
	(1.847)	
as.factor(exp)Sr		3.937***
		(1.489)
Constant	34.590	-28.281
	(45.147)	(55.029)
Observations	202	274
R2	0.168	0.124
Adjusted R2	0.133	0.098
Residual Std. Error	11.334 (df = 193)	11.982 (df = 265)
F Statistic	4.864*** (df = 8; 193)	4.702*** (df = 8; 265)
Note:		*p<0.1; **p<0.05; ***p<0.01

The results show a very large change in magnitude for both dummy variables. For players who declare as sophomores rather than as freshman, there is an expected difference of going five picks higher, or later, in the NBA draft, with strong statistical significance. For those who choose to stay in college for their senior season instead of leaving after their junior year, the players will see their draft selection drop by an average of four choices, which also had significant backing. This could be due to teams perceiving older players with more college coaching and less time to develop as being less useful to them unless they are close to a championship. We also see how there is little significance to any of the counting or efficiency stats for players in college, aside from box plus minus. However, it is nonetheless surprising that a higher offensive and defensive efficiency leads to higher, less desired draft selection in the draft following their final season. This could further the idea that teams draft more on potential than college production, but could also be related to bias that arises from not having other internal information teams collect such as private pre-draft workout stats, character assessments, etc.

The next model that was run was a regression discontinuity on pick using the mock draft consensus rankings. We are once again substituting the variable sr into the model for regression (2) in Table 2, which identifies seniors that are declared as the treatment group. The sensitivity around the bandwidth was very strong for both regressions, and due to complications with the rdrobust function, much of the analysis was handled manually. However, the study was conducted with 20% bandwidth on the regression (1) and a 10% bandwidth on regression (2) based on sample sizes and standard economic practices. The model and its results are as follows:

$$pick = \beta_1 porpag + \beta_2 dporpag + \beta_3 adjoe + \beta_4 adjde + \beta_5 bpm + \beta_6 ppg + \beta_7 rpg + \beta_8 apg + \beta_9 so + \beta_{10} highgrade + \beta_{11} so * highgrade + u$$

Table 2: Pick Regression Discontinuity

	Dependent variable: pick (1)	(2)
porpag	8.836*	-1.628
	(4.289)	(2.762)
dporpag	-8.811*	-1.060
	(4.504)	(2.767)
adj_oe	-0.754*	0.200
	(0.371)	(0.302)
adj_de	0.296	-0.031
	(0.465)	(0.338)
ppg	-2.958***	-0.090
	(0.772)	(0.357)
rpg	4.430***	0.269
	(1.122)	(0.420)
apg	5.793***	1.126
	(1.539)	(0.691)
as.factor(exp)So	6.921*	
	(3.622)	
as.factor(exp)Sr		-0.896
		(2.415)
high_grade	-7.262**	-18.631***
	(3.287)	(2.695)
as.factor(exp)So:high_grade	-0.096	
	(5.240)	
as.factor(exp)Sr:high_grade		5.155
		(3.274)
Constant	93.405	34.511
	(73.432)	(59.850)
Observations	32	164
R2	0.745	0.427
Adjusted R2	0.624	0.390
Residual Std. Error	6.492 (df = 21)	9.617 (df = 153)
F Statistic	6.138*** (df = 10; 21)	11.407*** (df = 10; 153)
Note:		*p<0.1; **p<0.05; ***p<0.01

The results show a strong effect of holding a high mock draft grade, as we see 7.262 boost in draft selection for all players in the freshman and sophomore regression and 18.631 expected increase in selection in the junior and senior regression. Both values are statistically significant at a a five percent alpha as well, so this would signal that this is an important feature. Using these results, this should lead all players who have a high mock draft grade, which with this cutoff was set at about the end of the first round, to declare for the NBA Draft as they are likely to be selected higher on average than their counterparts. This seems intuitive, but as has been discussed, there are many other factors that are not publicly available that could skew these findings. However, this is nonetheless telling that if the public data favors a player, it is seemingly optimal for them to declare for the draft. The interaction terms are also of note, as we see a very small boost in draft selection for highly graded sophomore players of about 0.1 spots, while we still see a similar relationship between waiting the extra year to declare as a senior in regression (2) as those that are highly graded are selected 5.155 picks later than highly graded junior counterparts. However, both terms are not statistically significant, so we cannot make any certain takeaways. We also see a strong significance on nearly all of the box score controls in regression (1), although it was interesting to note that higher rebounds and assists per game actually decrease selection number. This could in part just be based on the number of players at each position in the sample, as some positions could be more scarce than others, as well as a possible preference for certain types of players by teams. The r-squared value is much higher for regression (1), which intuitively makes sense as these players have less college experience and in turn these statistics and their draft grades will make up more of the information that teams have on players. Players in regression (2) who have been in college for numerous years likely have plenty of evaluation from the NBA, much of which is not available to the public. However, these results would seemingly support the alternative hypothesis that playing an additional year of college basketball does impact their selection in the NBA draft, although it is difficult to what level of significance.

The final regression run in our study was an OLS regression with the experience dummy variables on players career lifetime earnings. This was of interest as it takes into account players' long term payoff from college, as they hope to stick around the league and earn high wages. The setup of the model was the same as the past two, with the following formula substituting in seniors for regression (2). The results are as follows:

$$log(lifetimeearnings) = \beta_1 porpag + \beta_2 dporpag + \beta_3 adj - oe + \beta_4 adj - de + \beta_5 bpm$$
$$+ \beta_6 ppg + \beta_7 rpg + \beta_8 apg + \beta_9 sophomore + u$$

Table 3: Earnings Dummy Regression

	Dependent variable:	
	log(lifetime_earnings)	
	(1)	(2)
porpag	0.815**	1.194***
	(0.315)	(0.336)
dporpag	-0.357	-0.495
	(0.296)	(0.338)
as.factor(exp)So	-0.690***	
	(0.258)	
as.factor(exp)Sr		-0.217
		(0.207)
adj_oe	-0.056*	-0.097***
	(0.030)	(0.036)
adj_de	-0.070**	-0.072*
	(0.033)	(0.040)
ppg	0.050	-0.096**
	(0.043)	(0.046)
rpg	0.023	0.077
	(0.057)	(0.054)
apg	-0.050	-0.056
	(0.087)	(0.086)
Constant	26.661***	32.098***
	(5.546)	(7.025)
Observations	111	184
R2	0.159	0.104
Adjusted R2	0.093	0.063
Residual Std. Error	1.078 (df = 102)	1.348 (df = 175)
F Statistic	2.402** (df = 8; 102)	2.536** (df = 8; 175)
Note:		*p<0.1; **p<0.05; ***p<0.01

The coefficient on the sophomore dummy variable shows a 69.0% decrease for players who declare for the draft as sophomores when compared to their freshman counterparts. This value is significant at a 1% alpha level. For regression (2), we see an average 21.7% decrease in lifetime wages for players who choose to stay for senior season in comparison to those who declare for the draft as juniors. However, this effect is not statistically significant. These differences in wages are however in support of the alternative hypothesis that wages are impacted by an additional year of college, although we can only support this for certain for sophomores. We also see minor shifts in wages based on many counting stats that are significant, although the only major coefficient was PORPAG, as we see an increase in wages of 81.5% in regression (1) on average and 119.4% in regression (2). This makes sense intuitively that players who have a high impact in difficult college matchups will in turn earn more due to their strong influence on their team. These r-squared values are much smaller than that of the previous regressions, which in turn makes sense as earnings will be impacted by NBA performance and training, as well as other factors that occur after college that are not included in the model.

Conclusions/Directions for Further Research

After carrying out our analysis, we have revealed notable trends in the NBA draft selection and subsequent career earnings based on a player's decision on the timing of their declaration for the draft. We found significance in both dummy variables we tested, indicating a substantial impact on draft position based on years of experience. Players who chose to declare as sophomores experienced an increase of five picks in the draft (got chosen 5 picks later than freshmen), and players who chose to stay for their senior season experienced an average drop by 4 picks, both carrying statistical significance. This suggests an influence of the timing of draft declaration on teams' perceptions of players, which could be connected to the consideration of player development and usefulness of each player in reaching the goals set by the team.

While we did find significance in experience, we surprisingly did not find statistically significant evidence to suggest that college statistics carry an impact on draft selection. In fact, we found that higher offensive and defensive efficiency led to lower draft pick, which again may point to a desire for potential versus experience when choosing players for a team.

One limitation that needs to be discussed is the selection bias aspect when comparing mock drafts to team intelligence in the NBA draft process. Mock drafts, which are often based on expert opinions and analysis of public statistical information, may not always fully depict the decision-making process of an NBA team when selecting their next player. Mock drafts can be influenced by public perception, many forms of media, or other public opinions which can introduce bias into the player evaluations. Conversely, private team intelligence may involve internal assessments, private workouts with the team, and other confidential information that is not accessible to the public, so we could not use that information in our analysis. It is even rumored that players and their agents selectively disclose or withhold information about injuries or their personal preferences to create their own image and narrative to the public, so this is not always the best metric to measure which players will go to which team. The intentional misdirection can make this data confusing. Thus, selection bias emerges as players decide whether to enter the draft, when they decide to enter the draft, and teams choosing their players based on a combination of publicly available information and private insight. The mock draft is not a sole estimation tool to be used, there is a need for a more comprehensive understanding of influential factors in the NBA draft besides publicly available information.

Finally, our regression model found that juniors and seniors earn on average 26.89% and 30.4%, respectively, compared to freshmen who declare for the draft, sophomores experienced a decrease in earnings compared to freshmen. However, the statistical significance was inconclusive and may point to other confounding variables such as age or personal circumstance. Draft pick number did emerge as a significant factor, with a small expected decrease of 4.66% in career earnings for each lower pick, which would be expected, as we would expect teams to value players that they chose first in the draft over players who were chosen later in the draft.

To research this topic further, there are a few directions one could take. The influence of college coaching styles on NBA success could be studied, along with team programs, and conference competitiveness. The database we used, toRvik, has team and coaching statistics included, so that would be a good database to use to determine if these variables coils also be significant. We could analyze whether players from certain colleges or conferences show different patterns in terms of draft positions, starting salaries, and overall career outcomes. Additionally, you could extend the analysis beyond basketball and compare the impact of early entry decisions in college on professional outcomes across different sports, like football, baseball, and hockey, that have different rules for entering their respective drafts and a different number of athletes, with differing proportions of athletes who play professionally after college. Similarly, women's sports could be examined to determine if there are similar trends to what we found in men's basketball. We could also look to measure if the reason for earlier draft picks earning more could be a situation of sunk cost fallacy and identify if less productive, high draft capital players earn more due to their team's formerly perceived belief of their potential. With further time and resources, it would also be interesting to analyze international player outcomes, as these players enter the draft at many different ages and a variety of different leagues that have varied skill levels.

Bibliography

"Estimated Probability of Competing in Professional Athletics." *NCAA.Org*, www.ncaa.org/sports/2015/3/6/estimated-probability-of-competing-in-professional-athletics.aspx. Accessed 5 Dec. 2023.

Greene, Alexander C. "The Success of NBA Draft Picks: Can College Careers Predict NBA Winners?" *St. Cloud State University*, St. Cloud State University, 2015, pp. 1–86.

Groothuis, Peter A., et al. "Early entry in the NBA draft." *Journal of Sports Economics*, vol. 8, no. 3, 2007, pp. 223–243, https://doi.org/10.1177/1527002505281228.

Wolfe, Kevin Joseph. "An Analysis of the NBA Draft : Are Teams Drafting Better and Does College Experience Truly Matter." *North Dakota State University*, North Dakota State University of Agriculture and Applied Science, 2022, pp. 1–12.