University of Northern Iowa

UNI ScholarWorks

2019

Examining the NCAA selection process for bias against mid-majors

Jackson Twait
University of Northern Iowa

Let us know how access to this document benefits you

Follow this and additional works at: https://scholarworks.uni.edu/hpt

Recommended Citation

EXAMINING THE NCAA SELECTION PROCESS FOR BIAS AGAINST MID-MAJORS

A Thesis Submitted

In Partial Fulfillment

Of the Requirements of Designation

University Honors

Jackson Twait

University of Northern Iowa

December 2019

This Study by: Jackson Twait

Entitled: Examining the NCAA Men's Basketball Selection Process for Bias Against

    Mid-Majors

Has been approved as meeting the thesis or project requirements for the Designation

University Honors

16 December 2019   Dr. Bryce Kanago

16 December 2019   Dr. Jessica Moon

I.   Introduction

The NCAA (National Collegiate Athletic Association) Men's Basketball Tournament is a single-elimination style tournament that takes place every March in order to crown the best men's college basketball team in the country. Every year since 2011, 68 teams have made it through the grueling regular season for a chance at winning the title of best in the nation. The current system's 68 teams are selected by a committee of ten athletic directors of schools and commissioners of conferences, but only 36 teams are actually selected by the committee because 32 teams gain an automatic spot in the tournament by winning their conference tournaments. Most of the 36 teams, even though they do not have an automatic bid, know going into the tournament selection day that they will be in the tournament. It is the teams at the lower end of the 36, along with the fringe teams that do not make the tournament, that truly have their fate resting in the hands of the committee. Many of the teams that are in this block are considered to be mid-majors. However, this is not to say that the committee does not play a role in determining all of the teams' fates because the committee is also in charge of ranking all 68 teams in order from best to worst. This determines the schedule of each team for the tournament.

A mid-major is a team that is not a member of one of the "power-seven conferences." The power-seven conferences consist of the Big 10, Big 12, Big East, ACC (Atlantic Coastal Conferences), Pacific Coast Conference (Pac 12), SEC (Southeastern Conference), and AAC (American Athletic Conference). A team from outside these seven conferences has not won an NCAA tournament since 1990 which should shed some light on why one group of conferences is called mid-majors and the other is called high-majors (high-majors will be referred to as power-seven from here on out). Not only are mid-majors not winning the NCAA Tournament, they are also barely getting into the tournament. Of the 36 at-large bids in the 2019 tournament, four of

those teams were mid-majors. Along with their performance on the court, the power-seven teams have several other distinctions from mid-majors. One of which being they are also able to bring in a lot more revenue off of their basketball teams. All this said, is it possible that there is a bias towards the power seven that makes it more difficult for these mid-major teams to make the NCAA Tournament? In my paper, I will examine whether there could be a bias against these mid-major teams.

II.   **Literature Review**

BJ Coleman, Michael DuMond, and Allen Lynch review biases in selecting NCAA Tournament teams by using data on NCAA Tournament results over the past 10 years in the research paper "Evidence of Bias in NCAA Tournament Selection and Seeding" (2010). To investigate this question, the authors used a logit regression method and 41 different independent variables, such as number of neutral court wins and number of wins above .500 against teams in top 25 Rating Percentage Index (RPI). The RPI is a method the NCAA uses to rank college basketball by weighting wins and losses by the quality of teams played. So say if a team beats a bad team, their RPI ranking would rise a little, but if they beat a really good team, their RPI would rise by a substantial amount. The (binary) dependent variable in their model was whether or not the team received an at-large bid to the NCAA Tournament. A binary variable means, in this case, that if a team receives an at-large bid, the at-large data column for said team will contain a "1." If they do not receive an at-large bid, it will contain a "0." The same binary variable method will be used for mid-major as an independent variable as well. If the coefficient for mid-major is negative and the p-value is lower than .05, the data then implies that there is sufficient to reject the hypothesis that there is no evidence of a bias against mid-majors. If the

coefficient were to be positive with a p-value of less than .05, then the data would imply a bias towards mid-majors. The authors did, in fact, find evidence of a bias against mid-majors, and they also found "substantial evidence" that there is a bias towards teams with some type of committee representation. To explain this, there is a committee of people that are used to select the NCAA Tournament field. These people are usually either commissioners of a conference or an athletic director at a certain school. The main reason for this bias, they hypothesize, is money. Conferences that made the NCAA Tournament, in 2008, received $1,146,078 for each tournament game in which a team from their conference played. If this seems like a lot, to put it in perspective, the NCAA takes in more than a billion dollars off the tournament. If there are 67 games in the NCAA Tournament, the NCAA is paying out roughly 75 million dollars to schools. That is still less than 10% of a billion.

In 2015, Rodney Paul and Mark Wilson wrote "Political correctness, selection bias, and the NCAA Basketball Tournament" which is a research paper similar to the aforementioned article. It reviews whether or not there is selection bias towards high-major teams and teams with committee representation. They also make some criticisms of Coleman, DuMond, and Lynch's article stating that margin of victory should not have been left out as an independent variable, and their model suffers from multicollinearity due to it including both RPI and the Sagarian rating system. The Sagarian Rating System is a method of ranking teams based off of their win percentage but also weighting wins and losses based off the team that the win/loss is coming against. It also adds extra weight to road wins and takes margin of victory into account which is where it differs from RPI. Paul and Wilson use two different models. The difference being the rating system for the teams in the first model was RPI, and the other used the Sagarian Rating System. In the first, evidence of bias towards power-seven teams was found using a probit model

and whether or not a team received an at-large bid as the dependent variable; however, in the second, there was no bias towards power-seven teams found using the same model with one difference. The committee is supposed to rely mainly on RPI due to the Sagarian including margin of victory because the NCAA does not want teams running up the score on their opponents. By comparing these two models, they are also testing to see if the committee considers margin of victory when determining what teams make the Tournament.

Timothy Zimmer and Todd Kuethe (2007), in "Major Conference Bias and the NCAA Men's Basketball Tournament," did not investigate bias towards mid-majors, but rather, they tested for bias between certain power-seven conferences. To find this information, they used an ordinary least squares regression model with data from 1997-2006 NCAA Tournament games. Zimmer and Kuethe used this information to predict the dependent variable, score differential. If the score differential is positive, then the higher seed won. If it is negative, the lower seed won. This would be considered an upset and would mean the higher seed was overseeded and the lower seed was underseeded. If the score differential equals 0, then the game went to overtime. In their review, they found evidence of bias towards SEC schools and against ACC schools. More specifically, they found that SEC schools were overseeded by 2 seeds and ACC schools were underseeded by 2 seeds. However, I wish they would have looked into if the SEC commissioner or an athletic director had been on the selection committee during this time or rather, the opposite for the ACC. This meaning that a member from the ACC was not on the committee. I also would have liked for them to include mid-majors in their analysis rather than only focusing on teams from the power-seven.

III.   Method

In my analysis, I used the model used by BJ Coleman, Michael DuMond, and Allen Lynch to determine a bias against mid-majors in the NCAA Tournament. Their model used a logit analysis due to the binary nature of the dependent variable. The dependent variable in this case is whether or not a team received an at-large bid into the NCAA Tournament. There are 26 independent variables in my dataset. Since I was only looking for teams that had a chance at receiving an at-large bid, I confined my dataset to only teams ranging from the 20th best team in the Sagarian Rating to 80th best team. This is also what Coleman, DuMond, and Lynch did. The logic behind this is teams ranked 1-20, so far, have always made the tournament, and it is rare for a team lower than 80 to make the tournament. To differentiate my analysis from Coleman, DuMond, and Lynch's, and to include an idea introduced by Paul and Wilson, I chose not to include RPI. To repeat why, I am doing this because including RPI and the Sagarian Rating System would most likely lead to multicollinearity, as stated by Paul and Wilson. I am using Sagarian instead of RPI because Paul and Wilson already found evidence towards a bias against Mid-Majors using RPI, but they did not find evidence towards a bias using the Sagarian Rating System. I use more recent data to see if the bias still exists. In my dataset, I also removed any teams that had won their conference tournament meaning they had an automatic bid into the tournament, and I only included data from 2015 to the present due to major conference realignments taking place before then. In 2014, the Big East split into two conferences: the Big East and the American Athletic Conference. However, it wasn't until 2015 that the American Athletic Conference had established its foundation of teams, which it did primarily by pulling teams from mid-major conferences. This left the dataset at 266 teams with 108 receiving an at-large bid and 158 not. Of the 108 teams to receive an at-large bid, 16 were mid-majors (14.8%) meaning 92 teams were from one of the power-seven conferences (85.2%).

The data used in my analysis comes from basketball-reference.com and also from kenpom.com. My dataset varies a little from Coleman, DuMond, and Lynch's due to the difficulty of finding some of their variables, but also because I was able to use some variables they did not from kenpom.com. The independent variables included in the model are winning percentage which is a team's wins divided by the amount of games played (Win %), the Sagarian Rating System (SRS), strength of schedule which will have a higher number if the team plays a harder schedule, conference wins subtracted by conference losses is included because there is a difference in competitive balance between mid-majors and power-seven conferences which typically allows for mid-majors to win more conference games, home wins subtracted by home losses, away wins subtracted by away losses, points per game, points allowed per game, a binary number with a 1 if a team is a mid-major and a 0 if power-seven (Mid-Major), field goals made per game, field goal percentage (Field Goal %), three pointers made per game, three point percentage (3 Point %), free throws made per game, free throw percentage (Free Throw %), offensive rebounds per game, total rebounds per game, assists per game, steals per game, blocks per game, turnovers per game, personal fouls per game, whether a team had a member from its conference or its own athletic director on the NCAA Tournament selection committee (Selection Committee), the average height of the team measured in inches, the average age of the team's starters with a freshman equaling 1 up to senior being 4, and how good the team's bench players were.

Before I begin my analysis, one would logically expect a few variables to be statistically significant. For starters, one would expect win-loss percentage to be significant with a positive coefficient because the committee will look at how many games a team wins versus games it loses. Next, Sagarian rating system should be significant with a negative coefficient due to the

fact that the dataset is ordered by Sagarian rating system with the lower numbers being the better rankings. Strength of schedule is another variable one would expect to be significant with positive coefficient because the committee has said before that they pay a lot of attention to the teams that other teams beat and lose against. The old saying is "defense wins championships," so one would also expect points allowed per game to be significant with a negative coefficient. In Coleman, DuMond, and Lynch, they found conference games won versus conference games lost to be significant, so one would expect to find the same with a positive coefficient. As a reminder, if the coefficient on the binary variable for mid-major is statistically significant with a negative coefficient, it would imply that there is evidence towards the NCAA Tournament selection committee being biased against mid-majors, and if having a member on the selection committee is statistically significant with a positive coefficient, it would imply that they are biased towards teams with a representative on the selection committee.

Table 1 contains the data from the binary logit analysis. It includes the odds ratio, p-value, and standard error. Points per game and blocks per game were omitted due to multicollinearity. According to the analysis, there does not appear to be a bias against mid-majors when selecting teams for the NCAA Tournament. Not only does there not appear to be a bias against mid-majors, the results imply that there could be a bias towards mid-majors. The results read that a mid-major is 1.26 times more likely to receive an at-large bid than a power-seven team even after controlling for other variables. Also, there does not appear to be a significant bias towards teams with a member on the selection committee. The odds-ratio for having member affiliation on the selection committee is 2.99 meaning that a team with a member on the selection committee has nearly 3 times better chance of making the Tournament than a team without a member on the selection committee, but the p-value is only .227. This means the

hypothesis that the coefficient is zero cannot be rejected. Both of these results go against what Coleman, DuMond, and Lynch found in 2010. In their research, they used a dataset containing data from 1999 to 2008. Potentially, this means that the NCAA is doing a better job selecting who gets put on the selection committee, and it also indicates that mid-majors are no longer getting overlooked.

Variables that were significant at the 5% significance level were win loss percentage, Sagarian Rating System, points allowed per game, free throws made per game, three pointers made per game, field goals made per game, and turnovers per game. I expected the first three to be significant, but I was not expecting free throws made per game since they are only worth one point. Free throws, three pointers, and field goals all had positive coefficients, while Sagarian Rating System, points allowed, and turnovers all had negative. These are the expected signs for all these variables. For the Sagarian Rating System, every 1 unit of ranking that a team rises, i.e. goes from ranked 30th to 31st, that team would be 6.59 times less likely to make the NCAA Tournament. The rest of the variables' interpretations can be found in Table 1.

IV.   Conclusion

According to my analysis of the selection process of the teams competing in the NCAA Men's Basketball Tournament, there is no bias against mid-major basketball teams, and there is also no bias towards teams who have a representative on the selection committee. This could be either the school's athletic director or a conference's commissioner. While this is possibly due to a bias in the Sagarian Rating, it could also just be that the selection committee is losing its bias against mid-majors that was previously discovered by Coleman, DuMond, and Lynch and Paul and Wilson. Hopefully, the latter is the case because this means the mid-majors are finally

getting the respect they deserve from the selection committee. If I were to redo my analysis, similar to Paul and Wilson, I would create two models. One would include RPI as an independent variable, and the other would include the Sagarian Rating System. Another option would be to check for a bias within specific seeds for NCAA Tournament teams using another regression analysis.

Table 1

Independent Variable	Odds Ratio	P-Value	Standard Error
Mid-Major	1.2645	.824	1.3342
Selection Committee	2.9913	.227	2.7119
***SRS	6.5930	.002	3.9315
**Winning Percentage	6.200	.011	6.5900
***Points Allowed	5.6261	.003	3.3143
***Field Goals	.0222	.003	.0284
**Three Pointers	.2361	.035	.1616
**Free Throws	.2220	.017	.1395
*Steals	1.8868	.08	.6834
**Turnovers	.5081	.032	.1603

*** = Statistically significant at the .01 level

** = Statistically significant at the .05 level

* = Statistically significant at the .1 level

Work Cited

Coleman, BJ, Michael DuMond and Allen Lynch. 2010. "Evidence of Bias in NCAA

    Tournament Selection and Seeding." *Managerial and Decision Economics* 31, no. 7

    (October): 431-52.

    https://www.researchgate.net/publication/227650360_Evidence_of_bias_in_NCAA_tour

    nament_selection_and_seeding

"College Basketball Statistics and History: College Basketball at Sports." *Reference.com*, 2019,

    www.sports-reference.com/cbb/.

Paul, Rodney and Mark Wilson. 2015. "Political correctness, selection bias, and the NCAA

    Basketball Tournament." *Journal of Sports Economics* 16, no. 2: 201-13.

    http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.829.2094&rep=rep1&type=pdf

Pomeroy, Ken. "2020 Pomeroy College Basketball Ratings." *2019 Pomeroy College Basketball

    Ratings*, 2019, kenpom.com/.

Zimmer, Timothy and Todd Kuethe. 2007. "Major Conference Bias and the NCAA Men's

    Basketball Tournament." *Economics Bulletin* 12, no. 1 (Jan).

    https://experts.illinois.edu/en/publications/major-conference-bias-and-the-ncaa-mens-

    basketball-tournament