SESUG 2023 Paper 178

An Efficacy Rating for March Madness Tournament Seeding

Joe DeMaio and Nathalie Jones,
Kennesaw State University

ABSTRACT

In 1979 the NCAA men's basketball tournament (affectionately known as March Madness) began seeding teams with ranks 1 through 10 into four regional categories. Prior to this time, no ranking of teams was utilized in the tournament. In this paper we present a metric to measure the efficacy of team seedings by the selection committee as compared to tournament results. Utilizing this metric, we compute the efficacy of seeding for both men's and women's teams from 64 team tournaments. We then theorize and correlate potential influencing factors.

INTRODUCTION

Come March of every year, productivity by workers and students crawls to a drag as we glue ourselves to college basketball tournaments. For whatever reasons, we all have a favorite college team or three. However, when those teams leave a big (or small) dance, we must root for some other school. Frequently one seeks out an underdog as marquee teams already have enough zealots. Yes, I'm looking at you, Duke, and Chapel Hill.

Fans love a good Cinderella team. We wonder if the tournament selection committee loves them in the same way. After all, a victory by Cinderella potentially demonstrates inaccurate seeding judgement by the selection committee. Compared to tournament results, how accurately seeded are teams? Has seeding efficacy increased or decreased over time? What influencing variables exist? Does gender impact the efficacy rating of seeding teams?

First, we must develop a metric to measure the efficacy of team seedings by a tournament committee. The best efficacy value for this function comes from perfect seeding of teams. What does this look like? It would be a bracket where the higher seed (lower rank number) won every game until reaching the final four with only number one seeds remaining. Flipping those results provides the worst seeding where all four 16 seeds advance to the final four. Our metric sums the numerical seed value for each team in the tournament at a fixed round. Thus, $R(year, round(i), gender)$ is a multivariable function where *round*(i) is the sum of the seeds of victors in round i based on year and gender. The final rating function is $E(year, gender) = \sum_i R(year, round(i), gender)$. For a 64-team tournament with rankings 1-16, a perfect efficacy rating is $4(\sum_{i=1}^{8} i + \sum_{i=1}^{4} i + \sum_{i=1}^{2} i + \sum_{i=1}^{1} i) + 2 + 1 = 4(36 + 10 + 3 + 1) + 2 + 1 = 203$. Let's use this metric to compute $E(2023, m)$ and the extreme values of *E*.

Table 1. 2023 Men's College Basketball Results from NCAA.com

$R(2023, round(1), m) = 191$	$R(2023, round(2), m) = 78$	$R(2023, round(3), m) = 37$
$R(2023, round(4), m) = 23$	$R(2023, round(5), m) = 9$	$R(2023, round(6), m) = 4$

Table 2. Round scores for 2023 Men's Tournament

$R(2023, round(1), m) = 144$	$R(2023, round(2), m) = 40$	$R(2023, round(3), m) = 12$
$R(2023, round(4), m) = 4$	$R(2023, round(5), m) = 2$	$R(2023, round(6), m) = 1$

Table 3. Perfect Seeding round scores for 64 Team Tournament

$R(2023, round(1), m) = 400$	$R(2023, round(2), m) = 232$	$R(2023, round(3), m) = 124$
$R(2023, round(4), m) = 64$	$R(2023, round(5), m) = 32$	$R(2023, round(6), m) = 16$

Table 4. Worst Seeding round scores for 64 Team Tournament

At this point, our function $E(year, gender)$ ranges over the integers in the interval $[203, 868]$ where $E(2023, m) = 342$. Let's create a real valued function on $[0,1]$ with the transformation

$$e = \frac{E(year, gender) - 203}{868 - 203} = \frac{E(year, gender) - 203}{665}.$$

In terms of accuracy of seeding, smaller numbers are better than larger numbers. In a perfect seeding $e = 0$ and the worst seeding yields $e = 1$. We will call this value the error rate for a tournament seeding. In 2023, $e = \frac{342-203}{665}$=.201. We theorize that if the likelihood of any team beating any other team was 50% then e would be a linear function. That is certainly not the case in real life tournament play. Most values of e will occur closer to 0 than 1 given the experience and knowledge of the selection committee.

ABOUT THE DATA

In 1985, the NCAA tournament moved to the familiar 64 team field we are accustomed to today for men. The woman's tournament expanded to 64 teams in 1994. NCAA tournament bracket information, early entrant players, and NBA/WNBA player salaries were scraped from various sources either programmatically in Python or by hand when necessary and practical. Python libraries such as requests and beautifulsoup were used to pull NCAA tournament brackets and early entrant NBA players from Wikipedia. Other player information can be found through an already existing API: nba_api.

Webpages, like ESPN.com, do not allow web scrapers to call their websites' URLs for scraping purposes. These webpages can be manually collected into an excel spreadsheet. Data on WNBA salaries and early entrant players is very hard to find and is often scattered across multiple sources, making automated methods of collecting this information take more time than manually copying the data from the webpage. After collecting the data, it was then joined, analyzed, and visualized in SAS 9.4 using the SQL, sgplot, means, freq, and ttest processes while referencing the SAS documentation pages.

SUMMARY STATISTICS FOR 64 TEAM TOURNAMENTS FOR MEN AND WOMEN

League	Variable		N	Mean	Std Dev	Minimum	Median	Maximum
Men's	Year		38	2004	11	1985	2004	2023
	Yearly Error Rate		38	0.1482	0.0415	0.0391	0.1466	0.2331
	Winner's Seed		38	2.03	1.75	1.00	1.00	8.00
	Round	1	38	187	14	155	188	215
		2	38	72	12	49	73	94
		3	38	27	8	13	25	41
		4	38	12	5	4	11	26
		5	38	5	3	2	5	15
		6	38	2.03	1.75	1.00	1.00	8.00
	Sum of Rounds 1 - 6		38	302	28	229	301	358
Women's	Year		29	2008	9	1994	2008	2023
	Yearly Error Rate		29	0.0730	0.0240	0.0286	0.0722	0.1429
	Winner's Seed		29	1.34	0.67	1.00	1.00	3.00
	Round	1	29	166	9	150	166	192
		2	29	55	6	40	58	68
		3	29	19	5	12	20	28
		4	29	8	3	4	6	17
		5	29	3.34	1.34	2.00	3.00	7.00
		6	29	1.34	0.67	1.00	1.00	3.00
	Sum of Rounds 1 - 6		29	252	16	222	251	298

Table 5. Summary Statistics of e and $R(year, round(i), gender)$ for 64 Team Tournaments

Does the error rate differ based on gender. Figures 1 through 3 amply demonstrate so. Both distributions appear approximately normal with respective means close to medians. However, the maximum error for women is ever so slightly less than the mean and median error rate for men.

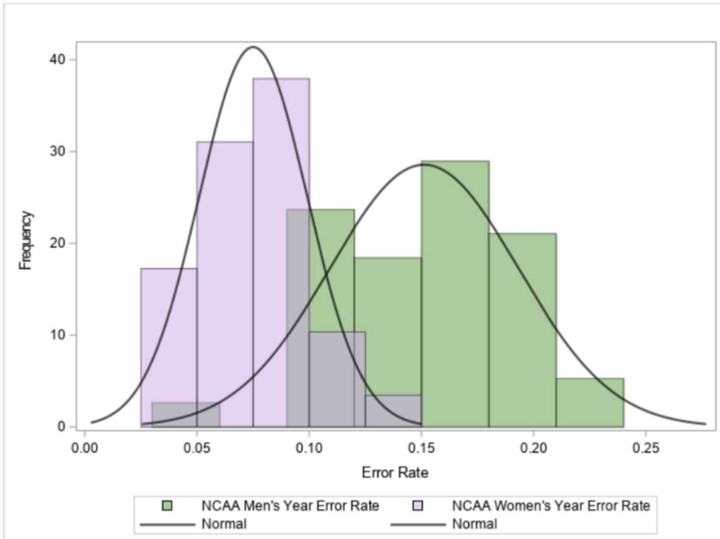

Figure 1: Distribution of Year Error Rates for NCAA Men's and Women's Leagues

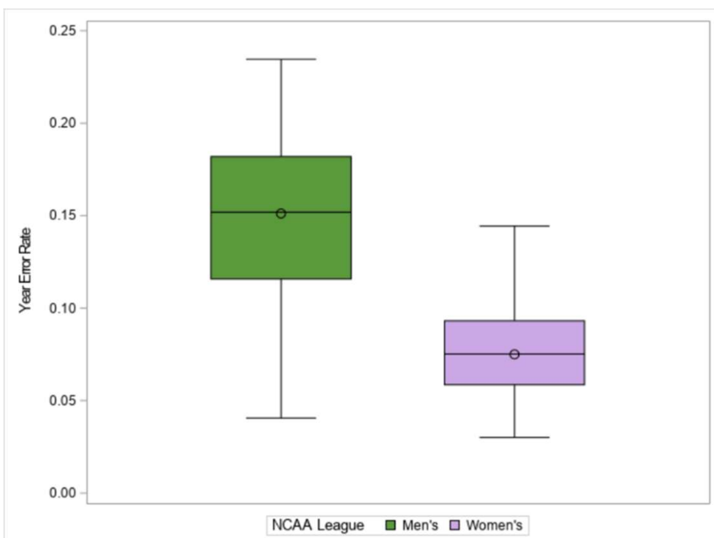

Figure 2: Year Error Rate for NCAA Men's and Women's Leagues

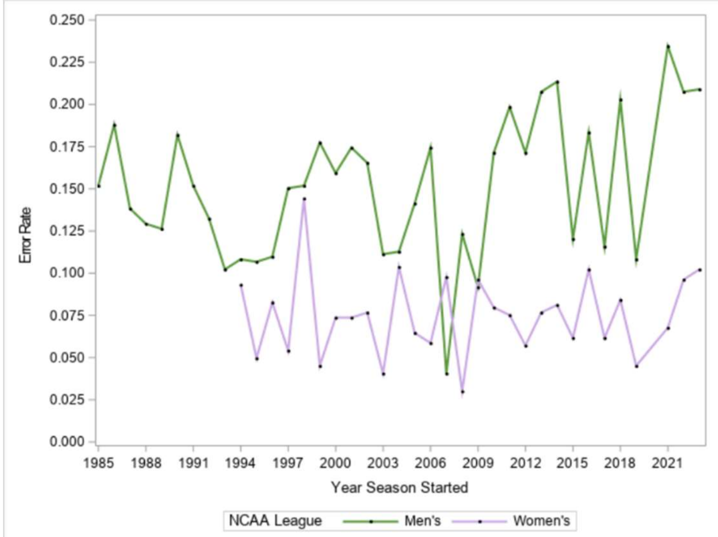

Figure 3: NCAA Seeding Error Rate for Men's and Women's Leagues Over Time

STATISTICALLY SIGNIFICANT DIFFERENCE

Given the visualizations of Figures 1-3, a natural question arises. Is the error rate significantly higher for men verses women? Let's conduct a one-tailed test at a 1% significance rate.

$$H_o\colon e_m = e_w$$

$$H_A\colon e_m \geq e_w$$

T-Test: Comparing Means of Error Rate Between Men's and Women's Basketball

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	65	8.74	< 0.0001
Satterthwaite	Unequal	60.906	9.36	< 0.0001

Table 6. Hypothesis Test for a difference of two means

Our tiny p-value indicates that it is much easier to seed women's teams versus men's teams. If one counts on NCAA seeding for assorted March Madness activities, the women's tournament is much more predictable. History has shown that in the men's tournament even a 16-1 matchup is unpredictable. Little Ceasars learned that the hard way by offering free pizza should a 16 seed beat a 1 seed. That happened when 16-seed University of Maryland Baltimore County beat 1-seed Virginia in the 2018 NCAA men's tournament in a 74-54 thrashing. In 2023 the men's 16-seed Fairleigh Dickinson beat 1-seed Purdue. Only once in women's play did a 16-seed beat a 1-seed when Harvard beat Stanford in 1998.

INFLUENCING FACTORS

It has been said that "money makes the world go 'round." Certainly, one must consider economics as an influencing factor. The NBA and WNBA pay their players at starkly different rates, with the highest paid 2023 NBA player, Steph Curry, making $51.9 million annually, and the highest paid WNBA players, Arike Ogunbowale, Jewell Loyd, and Diana Taurasi making only $234,936. Numbers for the WNBA start to look even more bleak for rookie players, with their salaries landing around $62,285 per year.

In the past 5 years, the average salary for a WNBA player was $97,381 while the average salary for an NBA player was $8,238,588. This may be in part due to each player's amount of time spent playing each season. The WNBA only hosts a 12-team league, and the number of games played each season is fewer than in the NBA. This means that it has less potential to bring in revenue compared to the NBA's 30-team league. Increasing the number of WNBA teams could potentially lead to an increase in play time for each player due to the higher number of games played in each season. In addition, the larger roster could accommodate more top ranked players each season and potentially bring more profit to the league.

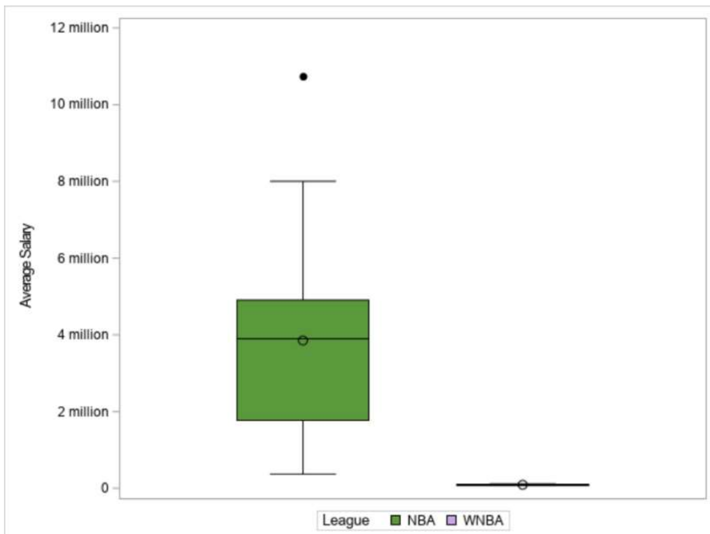

Figure 4: Average Salary for NBA and WNBA Players

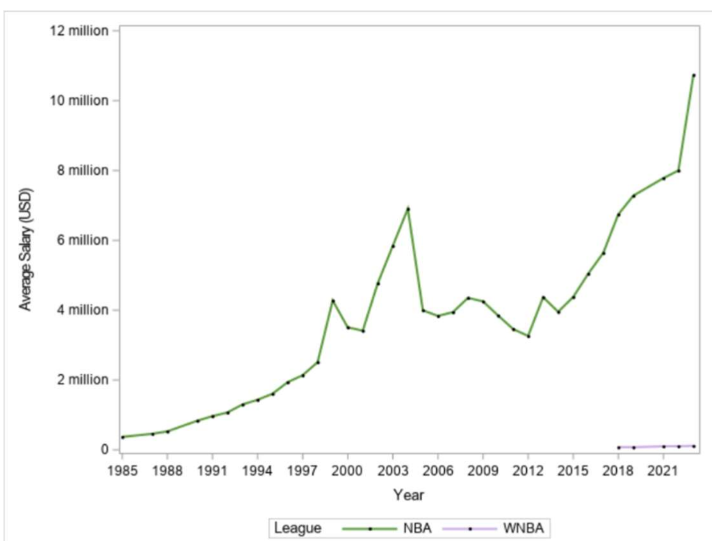

Figure 5: Average Salary for NBA and WNBA Players Over Time

Another example of the extremity of the pay gap between the two leagues reveals itself when looking at the highest paid WNBA players in 2022 (Diana Taurasi, Jewell Loyd, Breanna Stewart) who make only $228,094 per year, compared to the highest paid NBA mascots. Table 7 displays the top five highest paid NBA mascots in 2022. The top three mascots had a higher salary than the top three highest paid WNBA athletes. This disparity disappears when comparing the average salary of a WNBA player in 2022 ($114,470) to the average salary of a typical mascot ($60,000).

Team	Mascot	Salary
Denver Nuggets	Rocky The Mountain Lion	$ 625,000
Atlanta Hawks	Harry The Hawk	$ 600,000
Chicago Bulls	Benny The Bull	$ 400,000
Phoenix Suns	Go The Gorilla	$ 200,000
Charlotte Hornets	Hugo The Hornet	$ 100,000

Table 7. Top paid NBA mascots

WNBA players and non-players alike have spoken out about the disparity, with Liz Cambage voicing her concern over the fact that WNBA coaches make four times what the highest paid players do. For comparison, the highest paid NBA players for the past few seasons earned salaries greater than $40 million dollars, while the highest paid NBA coaches earn between $7 million and $13 million. In addition, eight WNBA players make less than $10,000 per year in return for their work.

We rely on this anecdotal information as making a direct comparison of NBA and WNBA salaries is challenging. ESPN freely provides NBA player salaries between 2011 and 2023 on their website. In contrast, salaries for WNBA players exist on one website, behind a paywall.

What impact might professional salary have on the efficacy of seeding March Madness tournaments? We conjecture that men are lured by astronomical salaries to the NBA before completing four years of college play. An early entrant player is defined as one who leaves school before completing four years.

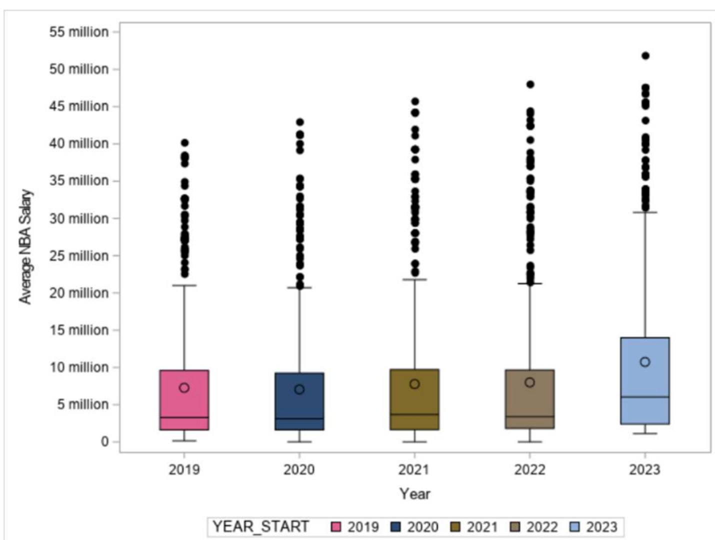

Figure 6: Average Salary for NBA and WNBA Players Between 1985 and 2023

Early entrants to the NBA and WNBA include players under the typical age of draft eligibility who displayed particularly excellent performance during their time on a college team. The

NBA wiki contains easily accessible data on early entrants for the NBA, but an analog does not exist for the WNBA.

According to the official WNBA rules and regulations, eligible domestic draft entrants must be at least 22 years of age during the draft year. For comparison, eligible NBA early entrants must be 19 years old or older and at least one year out of high school.

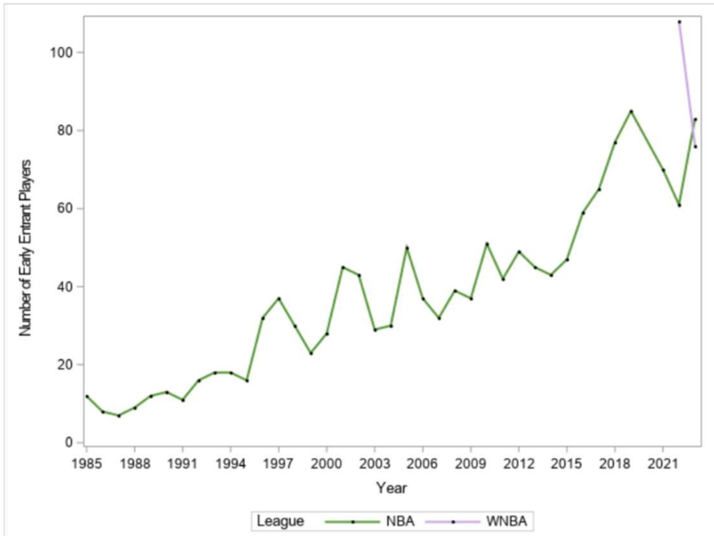

Figure 7: Number of Early Entrant Players into the NBA and WNBA Over Time

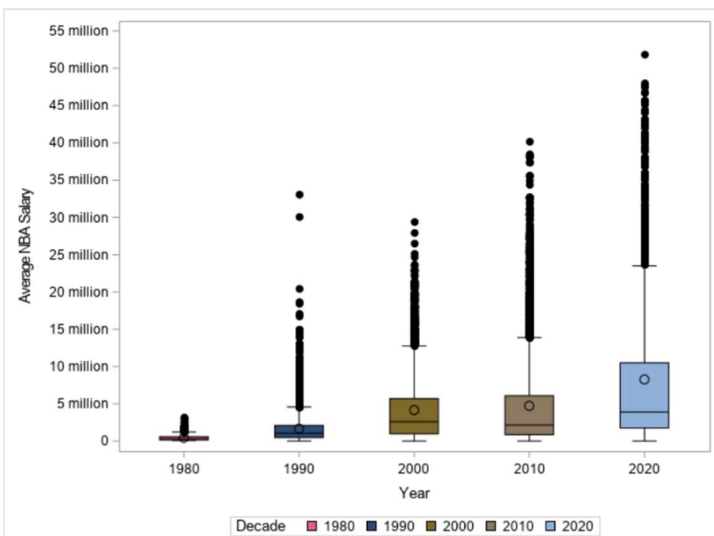

Figure 8: Average Salary for NBA Players by Decade

Average salaries for NBA players have increased since the men's tournament began seeding teams. Does this correlate to an increasing number of early entrant players?

The regression model predicting NBA player salaries suggests that for each passing year, the average salary for an NBA player should increase by an amount between $176,262 and $213,890, with an average increase of $195,076 per year (CI = 0.95, $p < 0.0001$***). The model also suggests that time accounts for 75.27% of the variation in average NBA incomes. The regression model predicting the number of early entrants into the NBA suggests that for each passing year, the number of early entrants should increase by 1 or 2 players, with an average of about 2 early entrants (CI = 0.95, p = 0.0001***). The model also suggests that time accounts for 85.52% of the variation in the number of early entrants admitted. Average salaries and the number of early entrants produce a correlation of 0.51 (p = 0.0011**), indicating a positive relationship. This suggests that as salaries increase, so should the number of early entrant players.

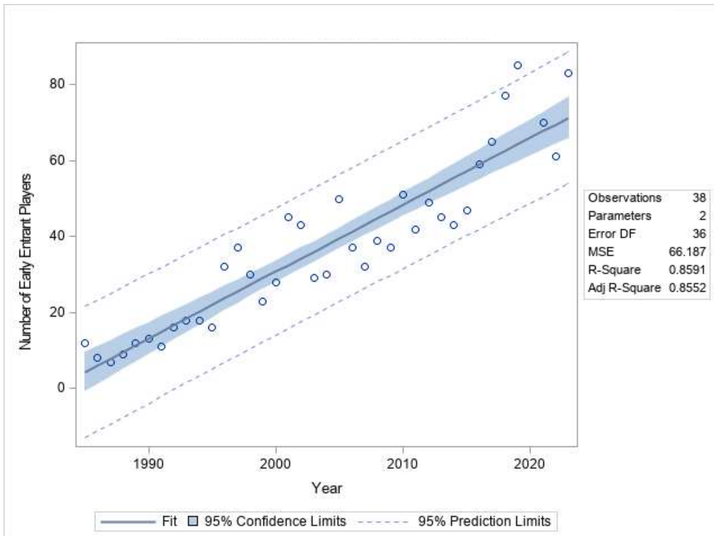

Figure 9: Fit Plot for Early Entrant Players

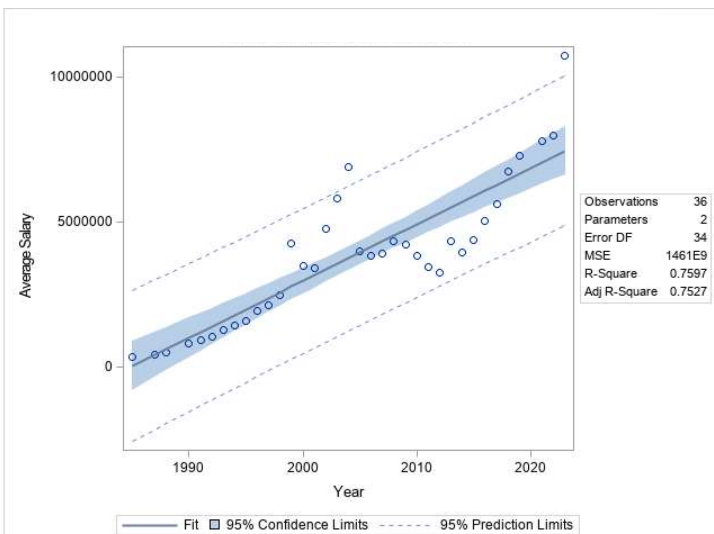

Figure 10: Fit Plot for Average NBA Salaries

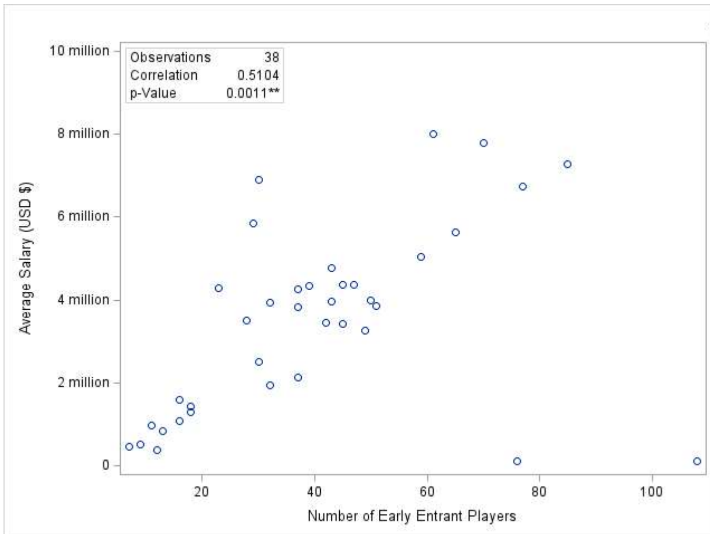

Figure 11: Correlation Plot for Average Salaries and Number of Early Entrant Players

FUTURE WORK

Earlier in this paper, we theorized that that if the likelihood of any team beating any other team was 50% then e would be a linear function. Future work consists of simulating tournaments to study the distribution of e relative to different probability estimators for winning teams. One such study will be conducted while fixing the probability of winning any given game at 50% for every team. Another such study will utilize historical probabilities from tournaments for the likelihood that an i seed beats a j seed for $i \neq j$.

For better comparison between the NBA and WNBA, future work will also consist of analyzing the amount of time spent playing for the top earners within each league, as well as the number of top performing players that did not get drafted. Comparing the time spent playing for top earners in the NBA vs WNBA could shed further light on the pay disparity between the two leagues. Additionally, the placement of top players during the draft could provide insights on the effect of the stricter drafting rules held by the WNBA and the disproportionately lower pay offered to their early entrant players.

CONCLUSION

With low error rates close 0, we conclude that the tournament selection committee does a good job seeding teams. However, seeding the men's tournament is much more challenging than seeding the women's tournament. We conjecture and provide some supporting evidence that economics plays a large role in this difference. We also find Cinderella teams exciting to watch and wouldn't have it any other way.

REFERENCES

**** NBA Draft. Accessed August 21, 2023. Available at [****NBA draft - Wikipedia] for **** running from 1985 To 2023

**** NBA Draft Early Entrants. Accessed August 21, 2023. Available at en.wikipedia.org/wiki/****_NBA_draft#Early_entrants for **** running from 1985 To 2023

Latest bracket, schedule and scores for the 2023 NCAA men's tournament. Accessed August 21, 2023. Available at [Latest bracket, schedule and scores for 2023 NCAA men's tournament | NCAA.com]

How to get a free lunch at Little Ceasars today for UMBC's historic win. Accessed September 1, 2023. Available at How to get free lunch at Little Caesars today for UMBC's historic win | Sporting News

NBA Player Salaries. Accessed August 21, 2023. Available at NBA Player Salaries - National Basketball Association - ESPN

NBA Salaries. Accessed August 21, 2023. Available at NBA Salaries | HoopsHype

NCAA Division I men's basketball tournament. Accessed August 21, 2023. Available at NCAA Division I men's basketball tournament - Wikipedia

NCAA Division I women's basketball tournament. Accessed August 21, 2023. Available at NCAA Division I women's basketball tournament - Wikipedia.

WNBA Draft 2023: Who are the early entrants, what are the eligibility rules, Accessed August 21, 2023. Available at WNBA Draft 2023: Who are the early entrants, what are the eligibility rules? - DraftKings Network

WNBA Salary Rankings. Accessed August 21, 2023. Available at WNBA Salary Rankings | Spotrac

ACKNOWLEDGMENTS

The authors would like to collectively acknowledge The School of Data Science and Analytics, a unit of The College of Computing and Software Engineering at Kennesaw State University, and the Analytics and Data Science student organization for travel support to present this work. We also thank and congratulate the 2023 Kennesaw Owls men for their first appearance in the March Madness Division I tournament.

Joe would like to thank Wake Forest athletics in general for teaching him to root for the underdog in basketball and football while simultaneously talking trash in golf. Specifically, he also thanks Tim Duncan, Randolph Childress and the rest of the '95 Demon Deacons for beating Carolina in the ACC tournament.

Nathalie would like to thank her wife Alexis for helping to brainstorm a data scraping plan, finding citations, proofreading, and brewing gallons of coffee. Her support through many late nights and research-engulfed days in an endless pursuit of data science contributed to the success of this project.

Nathalie would also like to thank Professor Susan Mathews Hardy, Professor Michael Frankel, and her other mentors from The School of Data Science and Analytics for providing her with the opportunity to learn and practice skills, such as statistical coding, web scraping, analytical writing, and presenting. Without their support and teachings, Nathalie's contributions to this paper would not have been possible.

Nathalie would also like to state that all ACC sniping contained in this paper is Joe's fault.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Joe DeMaio
Kennesaw State University
jdemaio@kennesaw.edu

Nathalie Jones
Kennesaw State University