

March Madness Model Meta-Analysis

What determines success in a March Madness model?

Ross Steinberg
Zane Latif

Senior Paper in Economics
Professor Gary Smith

Spring 2018



Pomona
College

Abstract

We conducted a meta-analysis of various systems that predict the outcome of games in the NCAA March Madness Tournament. Specifically, we examined the past four years of data from the NCAA seeding system, Nate Silver's FiveThirtyEight March Madness predictions, Jeff Sagarin's College Basketball Ratings, Sonny Moore's Computer Power Ratings, and Ken Pomeroy's College Basketball Ratings. Each ranking system assigns strength ratings of some variety to each team (i.e. seeding in the case of the NCAA and an overall team strength rating of some kind for the other four systems). Thus, every system should, theoretically, be able to predict the result of each individual game with some accuracy. This paper examined which ranking system best predicted the winner of each game based on the respective strength ratings of the two teams that were facing each other. We then used our findings to build a model that predicted this year's March Madness tournament.

Introduction

March Madness is an annual tournament put on by the National College Athletic Association (NCAA) to determine the best D1 college basketball team in the United States. There is one March Madness tournament for men and one for women. We focused on the men's tournament given the greater amount of analyses and model data available for it. The tournament consists of 68 teams that face each other in single-elimination games until there is only one team left. Out of all the teams, 32 receive automatic bids based on their regular season performance, and an NCAA selection committee chooses 36 additional teams. The NCAA selection committee seeds the teams based on perceived strength from 1 to 68, and eight teams play each other in

what are known as “play-in” games before the regular tournament. The four winners of the play-in games join the 60 other teams in the main March Madness tournament. Based on geographic region, these 64 teams are broken up into four groups of 16: Midwest, West, East, and South. Within each region, the teams’ playing schedules are arranged so that the highest seeded teams in each region play the teams with the respectively lowest seeds (e.g., seed 1 vs. seed 16 and seed 8 vs. seed 9). Interregional play occurs once four teams remain, also known as “The Final Four.”

In order to measure which system best predicts the results of March Madness, we first collected pre-tournament ranking data for each of the five ranking systems and the round-by-round results of the past four tournaments. We collected these data for the 68 teams in the tournament for the past four years. We then noted how many total wins each model correctly predicted for each year and aggregated each model’s total wins for all four years. After this, we ran a regression on the specific metrics included in the models to see which ones had the most predictive value in determining game outcomes. Using the results of this regression, we then created our own models that we tested on this year’s March Madness tournament.

More than just a way to determine the best DI college team, March Madness has become an important cultural touchstone. Each year, billions of dollars are bet on games throughout the tournament and millions of Americans fill out March Madness brackets. “Bracketology” has entered the common lexicon and during his tenure President Barack Obama filled out brackets in multiple years. A March Madness win can boost a college’s application numbers and help high school athletes determine to which college they wish to matriculate. In sum, there are a number of reasons people find it useful to have access to accurate predictions of March Madness

tournament games. This is why we analyzed the predictive values of both overall models and the individual metrics that go into the models.

Review of Literature

The literature review consists of an examination of two papers and how they relate to our meta-analysis. One (McCrea 2009) retroactively looked at how individuals predicted games in the NCAA March Madness tournament. It specifically focused on the factors people considered when they picked upsets and then evaluated the trends. The second paper (Toutkoushian 2011) discussed various prediction strategies and gave some background on these strategies, which provided context for the models we evaluated. This paper names common variables that many prediction systems use; however, it is backward looking and is focused solely on the creation of a model. Our paper in many ways serves as an update of their work, which was published in 2011. This is especially true as the paper did not reference some of the more popular prediction systems prevalent today.

The first paper looked at the ways in which individuals predict March Madness games. It found that people tend to predict upsets at a strategically nonoptimal rate; people were found to predict a number of upsets similar to the number of upsets that had occurred in the past. However, because individuals were unable to match their upset predictions with situations where upsets occurred, they predicted fewer games correctly than they would have if they had gone for the favorite team more frequently (McCrea 2009, p. 27). Further, the study found that even sports bettors who have adequate remunerative incentives to correctly predict games were not able to beat bookkeepers and that at an individual level expert predictions were no better than

non-expert predictions (McCrea 2009, p. 27). Despite the amount of luck involved in the tournament, this suggests that there is much progress to be made in better predicting March Madness.

The second paper shared a goal with our paper: develop a model that accurately predicts March Madness tournament game results. The paper used seed, data from the regular season (such as win percentage), and the average years in school of team players in order to create its model. It also looked at historical team data but found that they were not statistically significant in predicting wins and losses (Toutkoushian 2009, pp 29-32). The paper found that its model was more retrodictively accurate in determining game outcomes than predictively accurate when looking at the 2011 tournament (Toutkoushian 2009, pp 29-32). Our paper has drawn upon this by also focusing on data from the current regular season for our primary model, and it expands upon this by creating ancillary models that incorporate other predictive systems as well.

Meta-Analysis

As we mentioned in our introduction, we are analyzing five different NCAA Basketball ranking systems from the past four years in order to determine which system has the most predictive power. Over the past few years, some of these systems have changed the way in which they assign ratings or predict the winner of games; however, they have largely remained the same and the changes are minimal. In this meta-analysis, we provide a year-by-year breakdown of each model and give potential reasons for why certain models may have done better or worse. We also provide brief overviews for each of the years we are analyzing in order to contextualize

the results. In other words, 40 wins one year may not be equivalent to 40 wins the next year. Many of these models use similar factors as their inputs and are often linear combinations of one another. Note that for every year, there are 63 games played, not including play-in games. Thus, a prediction system could theoretically predict 63 games correctly, but there is no record of anyone ever creating perfect bracket. This section will give a brief overview of each prediction system, how they have changed (if at all), and provide year-by-year and total breakdowns.¹

NCAA Tournament Overview

2014 March Madness Tournament

In the 2014 March Madness Tournament, the higher seed won 40 of the 63 possible games. This represents the second most upsets in the four-year span we are analyzing. In conjunction with the upsets, every single computer-based system had its worst year in 2014.

2015 March Madness Tournament

In the 2015 March Madness Tournament, the higher seeded team won 50 of the 63 games. This tournament had the fewest upsets of any of the four years, and every prediction system we are analyzing had its best year in 2015.

2016 March Madness Tournament

The 2016 March Madness Tournament had the most upsets, but only by a margin of 1 game. This was the worst year for NCAA seeding and was seven games lower than the next

¹ Note that the description of the ranking systems and our methods for analyzing them are in the Appendix section.

worst prediction system. In fact, most of the difference between the NCAA and the other prediction systems over the four-year span is a result of this year.

2017 March Madness Tournament

For the 2017 March Madness Tournament, the higher seed won 46 of a possible 63 games. This is the second-highest total of the four-year span for NCAA seeding.

Meta-Analysis Results and Conclusions

In this section, we compare the ranking systems on a year-by-year and aggregate basis. The numbers under each prediction system are the number of games predicted correctly for the given year.

YoY Results

Below is a table displaying the year-by-year results:

	NCAA Seeding	Sonny Moore	FiveThirtyEight	KenPom	Jeff Sagarin
2014	40	44	42	39	46
2015	50	51	49.5	52	50
2016	39	47	46.5	50	48
2017	46	51	45	45	48
Total	175	193	183	186	192

A strong performance in the 2015 and 2017 March Madness Tournaments greatly assisted Sonny Moore's total. Moreover, the consistency of Jeff Sagarin's system helped him achieve a respectable total. Ken Pomeroy was clearly hurt by his poor 2014 March Madness

Tournament, but this year may be an outlier. The FiveThirtyEight's model's performance was consistently either middling or poor. Over the four-year span, all of the systems outperformed NCAA seeding by anywhere from 8 to 18 games.

The two most successful systems for the four-year span were Sonny Moore's Computer Power Rankings and Jeff Sagarin's "predictor" ratings. They correctly predicted 193 and 192 wins out of 252 games, respectively. We postulate that the reason for Jeff Sagarin and Sonny Moore's success over the other systems is due to the focus and purpose of their systems. While FiveThirtyEight aggregates different systems and Ken Pomeroy provides strength ratings, Sagarin and Moore's systems are specifically intended to be predictive. Because of this, they more heavily weight recent games, with Moore's systems putting the most emphasis on them. This focus on prediction rather than retrodiction may help explain why Sagarin and Moore are so effective. Moreover, both systems look at largely the same variables: margin of victory, wins, and losses. Indeed, more simple systems with only a few variables may tend to do better. (Note also that this covers only a four-year span due to data availability, so these results may change over time.)

Data

In figuring out which metrics to put into our model, we regressed a number of different variables to determine their retrodictive value in predicting previous March Madness games. We wanted to use our meta-analysis to inform our own model and chose metrics that were commonly used by other successful models. First, we used the "predictor" statistic from Jeff Sagarin, which is also used by 538 (Sagarin 2018). Next, we used NCAA seeding numbers,

which are also used by 538 and which we obtained by looking at past official NCAA brackets (“2018 DI Men's Basketball Bracket” 2018). We then used 538’s own Elo rating, Sonny Moore’s system--which 538 also incorporates into its model--and the Basketball Power Index (BPI), yet another one of the six computer rankings used in 538’s model (Silver 2017). We also used the win percentage of teams in the regular season, as this was the most significant and most often significant variable used by Wright (2012) for his economics senior paper. Note that we didn’t regress other factors found significant by Mr. Wright--such as opponent points per game, points per game allowed, etc.--as these are already used in Ken Pomeroy’s model, among others, and are thus implicitly included in the model; we chose not to include Mr. Wright’s paper in the review of literature section since we analyze it in the next section.

We also looked at Oddsshark betting spreads based upon the idea that, through the law of large numbers and the assumption of rationality, the market is an efficient way whereby to predict basketball games (“NCAAB Basketball Odds & Handicapping Database” 2018). Lastly, we looked at teams’ adjusted efficiency, which is the main variable that Ken Pomeroy used in his rankings, as well as adjusted defense and adjusted offense, which make up Mr. Pomeroy’s adjusted efficiency (Pomeroy 2018). For adjusted defense and betting spread, a lower number is associated with a better team; thus we expected these variables to have negative coefficients. For the rest of the metrics, a higher number is associated with a better team; thus we expected these variables to have positive coefficients. Note that we looked at the difference in value between a team’s metric and its opponent’s and regressed that difference. So, for example, when we used win percentage in our model, what we really used was the difference between a team’s win percentage and that of its opponent.

Model Creation and Regressions

Initially, we ran a logistic regression with the binary option of a team winning or losing being the dependent variable. The independent variables were Moore's power rating, win percentage, betting spread, Elo, predictor, seeding, BPI, adjusted efficiency, adjusted offense, and adjusted defense. Below is a table of our test results from Stata:

Variable	Coefficient	t-Score	P-Value	VIF
Seed	.018	.23	.819	3.48
Offense	-.154	-.21	.830	26.05
Defense	.192	.27	.789	14.73
Betting Spread	.130	.84	.401	8.36
Win %	8.208	2.53	.012	2.48
Predictor	.062	.28	.780	78.45
Elo	-.010	-2.00	.046	19.92
Moore	.491	3.30	.001	25.41
Efficiency	.172	.24	.812	65.31
BPI	-.078	-.44	.663	48.28
Constant	.093	.33	.742	Total: 29.25

As can be seen, only win percentage, Moore's power rating, and Elo were statistically significant at the 5% level; however when we found the variance inflation factor (VIF) of our regression, it was 29.25. VIF values above 20, or even 10, signify that the regression has a serious multicollinearity problem. We realized that when we used overall model metrics such as Elo, Moore, and adjusted efficiency along with the metrics used to create these models there

were multicollinearity problems. We had been modeling our paper off of Chris Wright’s paper, hoping to improve upon it with more recent and expansive data. However, Mr. Wright used both overall models and the metrics used to create these models in the same regression. This meant that there were significant overlaps in the independent variables of his regression, which led to coefficients being a different signage than they intuitively should have been (e.g., a team’s having a higher number of average points per game in away games had a statistically significant negative effect on the team’s chances of winning). Therefore, we pivoted to creating a simpler model to avoid multicollinearity, using only primary metrics rather than composite scores from the models themselves. We thus took out Moore, BPI, Predictor, Elo, and adjusted efficiency (which is adjusted defense subtracted from adjusted offense). In the new logistic regression, we included seeding, adjusted offense, adjusted defense, betting spread, and win percentage. The regression is below, and as can be seen our VIF is now only 2.85 since we have gotten rid of the overlapping variables.

Variable	Coefficient	t-Score	P-Value	VIF
Seed	.072	1.59	.111	3.59
Offense	.135	3.61	.000	3.57
Defense	-.100	2.69	.007	2.83
Betting Spread	-.059	-1.42	.156	2.45
Win %	1.88	1.30	.194	1.73
Constant	.065	.41	.682	Total: 2.85

As seen above, all of the variables have somewhat low p-values, with adjusted offense and adjusted defense being statistically significant at the 5% level. We decided not to include

betting spread. While the betting spread is useful in picking winners in games for which the opponents have already been chosen, a bracket must be filled out before the tournament begins. Thus, betting spread is not useful for our purposes as we want to try and create a successful March Madness bracket before any of the games take place.

Our main model thus used the statistically significant coefficients from the logistic regression that includes the variables above minus the betting spread. The new regression results are:

Variable	Coefficient	t-Score	P-Value	VIF
Seed	.054	1.25	.21	.30
Offense	.162	4.56	0.00	.43
Defense	-.125	-3.54	0.00	.48
Win %	1.99	1.36	.17	.58
Constant	.073	.16	.64	Total: 2.36

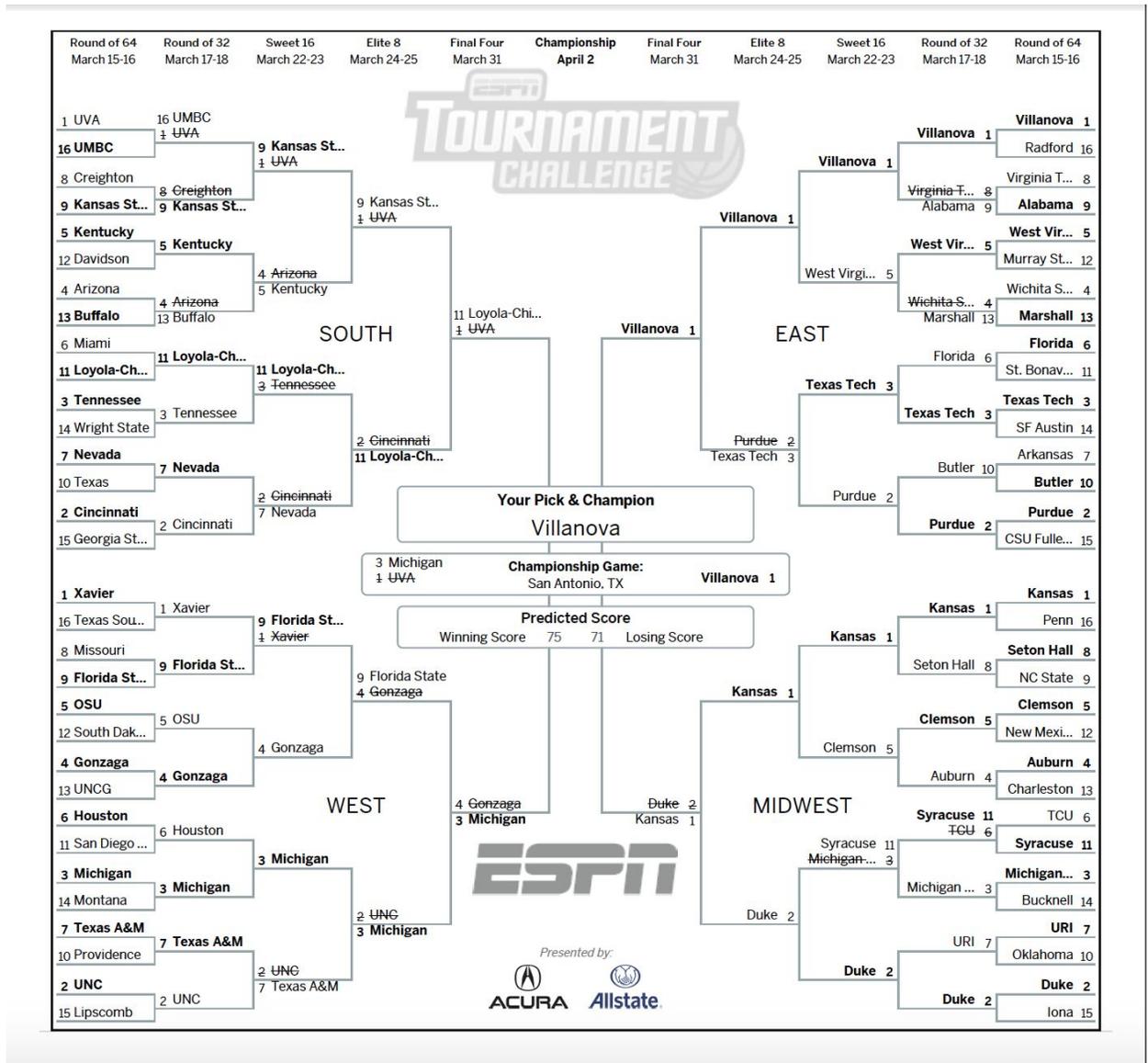
The only two variables still significant at the 5% level are adjusted offense and adjusted defense, but seeding and win percentage still have relatively low p-values. We created our main model by multiplying a team's adjusted offense by .162, its adjusted defense by -.125, its seed by .054, its win percentage by 1.99, and adding these four numbers together.² We then ordinally ranked the teams based on this score and predicted that in each matchup the team with the higher ranking (i.e., closer to one) would win. Although this is our primary model, we also chose to create three ancillary models. One of the models follows the same procedure as above, but only uses adjusted offense and adjusted defense, as these were the only two variables significant at the

² As explained above, a lower adjusted defense is associated with better defense.

5% level. This model also serves as a robustness check to ensure that our main model still obtains similar results with regressors taken out. We created a third model where we compiled five ordinal lists of team strength based on our primary model, FiveThirtyEight's model, Moore's model, Sagarin's model, and Pomeroy's model, averaged the five ordinal rankings together, and predicted that in each game the team with the highest ranking (i.e., closer to one) would win. Lastly, we followed the same procedure above using only our primary model and Moore's model, as his was the most successful model in the meta-analysis. By creating these extra models we aimed, in the latter two cases, to assess if our own purely metric-based model was better than a model that included the explicit predictions of other systems.

Model Results

Of the models we created, our primary one (wherein we used seed, adjusted offense, adjusted defense, and win percentage) was the most successful with 41 games predicted correctly. It was in the 95th percentile of all ESPN brackets created. Next successful were our simple adjusted offense / adjusted defense model and the multi-model average, both with 40 games predicted correctly. The worst performing model was the Moore model averaged with our main model, as it correctly predicted 39 games. As for the models created by others, FiveThirtyEight's predicted 39 wins in this upset-heavy year and Moore, Pomeroy, and Sagarin all predicted 38 wins correctly. Below is the image of our primary bracket, which also happened to be our most successful bracket:



The models, however, differed more in the earlier rounds than in the later ones. Each model correctly picked three of the teams that made it to the elite eight (Villanova, Duke, and Kansas) and one of the teams that made it to the final four (Villanova). Three out of the four picked Villanova as the champion while the model that used solely our regression and Moore's predictions picked Virginia as the winner. Although it is encouraging that our primary model still

predicted the most games correctly, the four models were all similar at the end of the tournament.³

It is promising that our main model and its variants out-performed or matched the success rate of the models in our meta-analysis; however, the closeness of the results and the fact that our own models were separated from each other by only a few wins makes it impossible to determine at this juncture if our main model is truly better or if luck was the differentiating factor. We will have to do similar tests in future March Madness tournaments in order to determine if our primary model is differentiated by factors other than luck.

Conclusion

In thinking about our results and the different aspects of our paper, we want to be critical of the inherent shortcomings of such an analysis. First, the models currently viewed as the best are thought to be so because they have done a good job of predicting games in the past few cycles. With so many people trying to predict March Madness games, it is very likely that a few individuals will create systems that correctly predict many games for a few years in a row; thus, it is hard to know at present whether these systems have true predictive power or merely seem so due to survivor bias. Additionally, not all the systems we analyzed are exclusively predictive in purpose. For instance, Ken Pomeroy assigns adjusted efficiency ratings and uses this as a proxy for a predictive rating, but that is not the same as Jeff Sagarin's "predictor" rating.

However, this is a somewhat moot point for people using these systems to make predictions, as they can only use the data publicly available to them. It may be interesting to see

³ Sites such as ESPN award more points for wins predicted at the latter stages of the tournament.

what results he would produce should Ken Pomeroy decide to reformulate his rankings into an explicitly predictive algorithm. These models are also all extremely similar--save for the NCAA model--because they use very similar variables and often borrow from one another. This makes analyzing them a trying task. Moreover, because they are very similar and often use the same variables, running regressions on the components of the models leads to multicollinearity problems.

In thinking about the difficulties regarding the construction of our own models, we understood that perfect accuracy is nearly impossible. There is such a great deal of luck and randomness that hoping for this is futile. In fact, there is no record of a perfect bracket. There is also no consensus method for constructing the best prediction system or bracket. While one tournament is not enough to determine the success of our models, the tournament results lend some insight into how well relatively simple models can perform in March Madness.

Our primary takeaways from our paper range from meta-analytic observations to implications on the creation of future models. With regard to our meta-analysis of the different models, we observed that, at least recently, models that are specifically intended to predict rather than assign strength rankings tend to perform better. This may be because they weight recent games more heavily and use simpler, more straightforward metrics. The two more heavily predictive models we analyzed mainly took into account margin of victory, wins, and losses; they also adjusted for strength of schedule and pace of play.

As we mentioned above, we also came to the conclusion that creating models for March Madness is a difficult task. More than this, it is and will always be virtually impossible to create a perfect model due to the inherent luck involved in college basketball. This year the primary

model we created performed better than the models we analyzed in our meta-analysis, but this does not guarantee future success. Indeed, just this year a number one seed, the University of Virginia, lost to a 16-seed, an historic and unlikely outcome that few models would ever predict. We would need perhaps an infinite number of seasons and iterations to come up with an optimal model. A last important takeaway based on our analysis is that if many of these models are similar, the popularity of these systems may be due to factors other than exclusively their historical predictive success. Prediction systems from experts or analytics gurus may be popular due to the media circus around them and their historical significance as much as for their predictive power or theoretical basis. Future research could still be done on how to better analyze, differentiate, and improve upon current popular models.

Appendix

NCAA Seeding

Overview

The NCAA is drastically different from any of the other four prediction systems. For one, their rankings are human generated as opposed to computer generated. While statistics are taken into account, they are not the end-all be-all for the overall seeding. While we will not go into as much detail on the March Madness Tournament as a whole in comparison to our introduction, some requisite background knowledge may be useful. The NCAA selects 68 teams overall (8 of which compete in play-in games in order to widdle the field down to 63). The NCAA ranks these teams from 1 to 68, which is known as the 68-team S-curve. The tournament then proceeds in a single-elimination fashion until there is one winner.

Team seeding is a complex and subjective process. While we will not describe this in excruciating detail, it can be broken into three essential parts:

1. Selecting the teams
2. Seeding the teams
3. Placing the teams in the regional brackets

32 teams receive an automatic bid after they win their conference championship and the other 36 teams are selected by the selection committee. This committee is comprised of athletic directors and conference commissioners in Division I men's basketball. Essentially, this selection committee votes by secret ballot in order to determine the 36 “at-large” teams. A similar process is used for the overall seeding. Note that in our meta-analysis, where a team is placed regionally does not affect the results. The important aspect of the NCAA seeding is that human beings

determine these rankings. The rankings are then revealed in an event called Selection Sunday. One of the purposes of our meta-analysis was to determine if the computer-generated systems perform better than the human-generated system of NCAA seeding.

Methods

We performed an analysis and awarded points on the following basis:

- We awarded 1 point for a negative difference (indicating higher seed or a higher ranking on the 68-team S-curve if the same seed) and a 1 in the “win / loss” column
- We awarded 1 point for a positive difference (indicating lower seed or a lower ranking on the 68-team S-curve if the same seed) and a 0 in the “win / loss” column

We then aggregated year-by-year and total results for NCAA seeding data.

Ken Pomeroy

Overview

The first rating system we are analyzing is Ken Pomeroy’s College Basketball Ratings. Pomeroy, who previously worked full-time at the National Weather Service as a meteorologist, is now perhaps the most well known college basketball statistics analyst. His predictions and advanced statistics are used by casual fans, betting markets, and even NCAA and NBA coaches and staffs in order to construct game plans.

His rating system recently underwent some minor changes in 2016, but fortunately, he updated all of the previous years to remain consistent with the new format. However, he did not

update his predictions, only his team rankings. The most notable change is that Pomeroy used to filter his projections through the Pythagorean Winning Percentage Formula, initially developed by Bill James for baseball (and referred to as Pythagorean expectation). This formula, which NBA executive Daryl Morey modified for use in basketball, is as follows:

$$\text{Winning percentage} = (\text{Points for}^{13.91}) / (\text{Points for}^{13.91} + \text{Points against}^{13.91})$$

Pomeroy further modified this formula and it read as follows:

$$\text{Pythagorean winning percentage} = (\text{AdjO}^{11.5}) / (\text{AdjO}^{11.5} + \text{AdjD}^{11.5})$$

Pomeroy decided to change his methodology for two primary reasons. First, in his own words, “it was a mouthful.” Second, it made comparing team strengths very challenging, as he states with this example: “The difference between .98 and .97 is not the same as the difference between .52 and .51 in terms of team strength.” His new formula is more user friendly, but in his estimation equally or more effective than his old methodology. Pomeroy now uses AdjEM - also known as adjusted efficiency margin - which is the difference between a teams offensive and defensive efficiency. That is, his new formula is just the subtraction of two different metrics. Adjusted offensive efficiency (AdjO) represents the number of points a team is expected to score when adjusted for strength of schedule and tempo. In other words, the number of points a given team would be expected to score against the average Division 1 opponent over a 100 possession span. A high offensive efficiency represents a team that is very good offensively. Adjusted defensive efficiency (AdjD) represents the number of points allowed over a 100 possessions

span, adjusted for strength of schedule. Thus, a low AdjD is indicative of a team with very good defense (Pomeroy 2018).

Methods

In conducting our analysis of Ken Pomeroy's College Basketball Rating systems, we collected ratings data from Ken Pomeroy's website for the past four years. We then created data columns in which one of two teams was listed and the other team's data was inputted as the "Opponent" column. A column for win or loss was added as well, with an input of a 1 representing a win, and an input of a 0 representing a loss. We proceeded to calculate the difference in adjusted efficiency margin between the two teams. We performed a binary analysis, awarding a point for either a positive difference and a 1 in the "Win/Loss" column or a negative difference and a 0 in that respective column. After inputting these formulas, we then aggregated year-by-year and total results.

FiveThirtyEight

Overview

FiveThirtyEight is markedly different from the other ratings systems we are analyzing for a number of reasons. First and foremost, FiveThirtyEight does not exclusively deal with sports analytics. They look at a variety of topics from politics to sports to economics and more. They generally try to provide a more rigorously statistical approach to their analyses than most and are perhaps best known for their analysis of presidential elections.

For their March Madness predictions, FiveThirtyEight uses a significantly different methodology from the other ranking systems we are examining. Unlike the other systems, which either utilize team statistics to come up with their own ratings or use team performance to rank them, FiveThirtyEight uses data from the other models alone in creating its own model. Specifically, they average a number of models (some computer generated and some human generated) and equally weight each model in coming up with their predictions. They then adjust for injuries and how far a team has to travel. To account for injuries, they use individual player win shares to estimate their overall impact on the team.

In 2014 and 2015, the FiveThirtyEight Model used five computer generated models.

Specifically, they included:

- Ken Pomeroy's College Basketball Ratings
- Jeff Sagarin's Predictor Ratings
- Sonny Moore's Computer Power Rankings
- Joel Sokol's LRMC Ratings
- ESPN's Basketball Power Index (BPI)

They note that all of these systems have decent predictive power and are generally composed with the same information such as wins, margin of victory, strength of schedule, etc. They claim that because the systems are all good and similar they are justified in weighting them equally.

Moreover, according to FiveThirtyEight, using multiple systems balances out any abnormalities in a single system. FiveThirtyEight also used two human generated ranking systems:

- The NCAA selection committee’s 68 team “S-curve” (i.e. the NCAA team rankings)
- Preseason Rankings from the Associated Press and the Coaches Poll (both preseason rankings are equally weighted)

In 2016, FiveThirtyEight added a sixth computer ranking to their formula. This computer ranking is called Elo Ratings, and they calculated it themselves. FiveThirtyEight used Elo systems for other sports previously, and they finally decided to bring it to college basketball. Elo is calculated by taking into account only the final score of games, home-court advantage, and the location of each game (i.e. distance travelled to get to a game). Thus, for 2016 and 2017, FiveThirtyEight equally weighted six computer generated systems, instead of five the previous two years, along with the two human generated systems (Silver 2018).

Methods

In conducting our analysis of FiveThirtyEight’s March Madness Predictions, we collected round-by-round predictions for the last four years from the FiveThirtyEight website. We then created a column that had the team name and another column if it was favored at or above a 50% chance to win. We did this for each of the six rounds in each of the tournaments. A column for win or loss was added as well, with an input of a 1 representing a win, and an input of a 0 representing a loss. We awarded points on the following basis with a winner take all system unless the prediction was dead even:

- .5 points for teams that were given a 50% chance to win
- 1 point for a team that was given a greater than 50% chance to win and had a 1 in the win / loss column
- 1 point for a team that was given a lower than 50% chance to win and had a 0 in the win / loss column

After conducting the analysis, we then aggregated year-by-year and total results.

Jeff Sagarin

Overview

Jeff Sagarin is often considered the pioneer of NCAA College Basketball analytics. He was one of the first individuals to come up with an individual number for a team rating. In fact, Sagarin's ratings are often used on their own as a variable in many models or prediction systems (e.g. FiveThirtyEight's predictive model). Jeff Sagarin has many different team rating scores, but the most used and referenced one is his "predictor" rating because it is meant to be the best predictor of upcoming games rather than simply a team strength score. Sagarin received a Bachelor's of Science in mathematics from the Massachusetts Institute of Technology, and with a former classmate, currently advises the Dallas Mavericks on lineup selection and free agent acquisition strategy.

Like many other NCAA College Basketball statistics analysts, Sagarin does not reveal exactly how he constructs his system. However based on how similar systems are constructed and the information he does divulge, there is enough to analyze his system. Moreover, because his system is primarily used as a single variable, the inner workings of his algorithm are less

important than the value of his ratings predictive power. Jeff Sagarin's "predictor" system is essentially based on the theory that score is the only thing that matters in predicting games. His "predictor" system, like many of the computer rankings, takes victory margin, wins, and losses into account. The difference in the "predictor" rating is intended to predict the margin of victory. Strength of schedule is taken account, as teams that win games against strong teams are awarded more points. Home-court advantage is also factored in, as teams that win away from home are awarded more points as well (and vice-versa) (Sagarin 2018).

Methods

Our methods for conducting our meta-analysis of Jeff Sagarin's "predictor" ratings were similar to that of other rating systems, such as Ken Pomeroy's College Basketball Ratings. In conducting our analysis of Jeff Sagarin's Predictor Rating system, we collected predictor variable data from Jeff Sagarin's website for the past four years. We then created data columns in which one of two teams was listed and the other team's data was inputted as the "Opponent" column. A column for win or loss was added as well, with an input of a 1 representing a win, and an input of a 0 representing a loss. We proceeded to calculate the difference in adjusted efficiency margin between the two teams. We performed a binary analysis, awarding a point for either a positive difference and a 1 in the "Win/Loss" column or a negative difference and a 0 in that respective column. After inputting these formulas, we then aggregated year-by-year and total results.

Sonny Moore

Overview

Like many of the other prognosticators, Sonny Moore began his Computer Power Rankings as a hobby in 1974 so that he could compare any two teams in a given sport. His system is specifically created to have predictive intention, as opposed to other systems that may be more oriented towards overall strength ratings or aggregating accomplishments (i.e. seeding). His system is meant to predict the number of points a team would win by should they play one another. He always had an interest in objectively determining the rankings of teams in a given sport from best to worst, and consequently, he devised his current system.

Sonny Moore's Computer Power Rankings are often considered very similar to Jeff Sagarin's "predictor" rankings because they also primarily take into account wins, losses, and scoring margin, and are adjusted for strength of schedule. Moore, also similar to Sagarin, applies a diminishing returns principle to his rankings so running up the score against a weak team does not overly reward a team. Sonny Moore's Computer Power Rankings do deviate from Sagarin's "predictor" ratings in one very important aspect: they more heavily weight recent games. In fact, Moore is known for creating a system that heavily weights recent games more than that of other similar systems (Moore 2018).

Methods

Our methods for conducting our meta-analysis of Sonny Moore's Computer Power Rankings were similar to that of Jeff Sagarin's "predictor" ratings and Ken Pomeroy's College Basketball Ratings. In conducting our analysis of Sonny Moore's Computer Power Rankings

system, we collected the data in the “PR” column on Sonny Moore’s website (in the archive section) for the past four years (2014-2017). The “PR” data refers to his team strength ratings.

We then aggregated this data in columns that matched with one of the teams in each of the rounds for the four years. We then aggregated the “PR” data for the opponents of these teams. After completing these steps, we performed a binary analysis. We calculated the difference between the two teams and added a “win / loss” column where a 1 represented a win and 0 represented a loss. We then constructed formulas to award points on the following basis:

- 1 point is awarded if there is a positive difference and a 1 in the “win / loss” column
- 1 point is awarded if there is a negative difference and 0 in the “win / loss column

After inputting these formulas, we then aggregated year-by-year and total results.

Bibliography

“2018 DI Men's Basketball Bracket.” *NCAA.com*, 2018,

www.ncaa.com/interactive-bracket/basketball-men/d1.

McCrea, Sean M., and Edward R. Hirt. “Match Madness: Probability Matching in Prediction of the NCAA Basketball Tournament¹.” *Journal of Applied Social Psychology*, vol. 39, no. 12, 2009, pp. 2809–2839., doi:10.1111/j.1559-1816.2009.00551.x.

Moore, Sonny. “Sonny Moore's Computer Rankings.” *College Basketball Power Rating*, sonnymoorepowerratings.com/m-basket.htm.

“NCAAB Basketball Odds & Handicapping Database.” *Odds Shark*, 2018,

www.oddsshark.com/ncaab/database.

Pomeroy, Ken. “2018 Pomeroy College Basketball Ratings.” *2018 Pomeroy College Basketball Ratings*, 2018, kenpom.com/.

Sagarin, Jeff. *College Basketball Ratings Page*, 2018, sagarin.com/sports/cbsend.htm.

Silver, Nate. “It's Crowded At The Top Of The NCAA Tournament.” *FiveThirtyEight*, FiveThirtyEight, 17 Mar. 2017,

fivethirtyeight.com/features/its-crowded-at-the-top-of-the-ncaa-tournament/.

Silver, Nate. "2018 March Madness Predictions." *FiveThirtyEight*, 3 Apr. 2018,
projects.fivethirtyeight.com/2018-march-madness-predictions/.

Toutkoushian, Emily. "Predicting March Madness: a Statistical Evaluation of the Mens
NCAA Basketball Tournament." *Ohio State University*, 2011.

Wright, Chris. "Statistical Predictors of March Madness: An Examination of the NCAA
Men's' Basketball Championship ." *Pomona College*, 2012.