Is there a Method to March Madness?

Jarrett S. Karnibad

Bolte School of Business

Mount St. Mary's University

Abstract:

Each year, millions and millions of people fill out brackets, and compete against their friends, families, coworkers, celebrities, and others around the world to try and guess the winners of each game in the tournament. And undeniably, each year, millions and millions of people fail. They watch as a team they had chosen to make a deep run in the tournament, lose in the first round to a largely unheard-of team from a small conference. Because March Madness is undoubtedly a fan-favorite sporting event, over the last few years, many have tried to develop formulas and criterion by which to determine the winners of NCAA tournament games. Each attempt has proven unsuccessful. This past year however was different. Through testing of regular season games and conference tournaments, two formulas were developed that showed a lot of promise: XPoint and KTheory. Both self-designed formulas give an expected final score of the game based on a multitude of different factors. This project examines the results of both formulas for the 2022 March Madness basketball tournament and ponders the question if it is possible to strictly use statistics to help provide some guidance to one of the most unpredictable sporting events in the world.

Disclaimer: All bets placed in this study are imaginary and are used simply to help provide another layer of analysis to the two formulas: XPoint and KTheory.

1) Introduction:

In American sports, there are few sporting events that truly capture the attention of the full public. The Super Bowl is one of them; the Olympics are another. But there is one event that causes people, who have no interest in sports otherwise, to care, even if for only a couple days. This is the NCAA basketball tournament, otherwise known as March Madness. There is a long and rich history to the NCAA basketball tournament, and it is one of the most popular sporting events in the United States. People who do not even like or watch basketball fill out brackets to compete in family or office pools. Year after year, the NCAA basketball tournament is one of the most watched sporting events of the year in the United States. It is easy to see why the tournament garners so many views. The amazing stories of teams and fanbases, the unpredictability of every game, and allure of watching the teams you picked attempt carry out your "perfect brackets" are all powerful reasons that people across the country tune into the broadcasts. For a majority of the country, the March Madness tournament is the one month out of the year, that they will on college basketball with intense focus and passion, either rooting for the alma mater or the teams they selected in their bracket. However, for a small minority of the country, the tournament provides an opportunity to accomplish the impossible: "creating a perfect bracket." This small minority studies the games at length, looking at teams' wins and losses, strength of schedule, and a multitude of power rankings. My two formulas use a combination of statistics in order to try and determine the final score of any given game. XPoint determines the final score through teams shooting percentages and average opponents shots allowed. This gives an estimate of how many points a team will score against another team's defense. If Team A allows 20 3-pointers per game, and Team B makes 50% of their 3-pointers, we can estimate that Team B will make 10 3-pointers, for 30 points, against Team A. The same

goes for 2-pointers and free throws. The other formula: KTheory attempts to predict the final score through a points per possession format. Much like XPoint, KTheory gives the expected final score for any game. The first key component is finding the effective possession ratio. This reflects the amount of scoring opportunities per possession. The second component of this formula is the True Shooting Formula. This represents the amount of points-per-shot opportunity. When multiplied together, the effective possession ratio and the true shooting formula give the points per possession. When multiplied by a team's average possessions per game, that will give the team's expected points per game. These two formulas are both entirely self-created and all data input was completed personally for every game for the tournament.

2) Related Literature:

This is not the first attempt to try to use statistical analysis to predict the winner of the March Madness tournament. In fact, there was a peer reviewed research article written by a group of authors: Lo-Hua Yuan, Anthony Liu, Alec Yeh, Aaron Kaufman, Andrew Reece, Peter Bull, Alex Franks, Sherrie Wang, Dmitri Illushin and Luke Bornn. They pose a very similar question that this study is trying to answer: Can you use statistics and power metrics to correctly predict the outcome for the NCAA March Madness tournament? This article included more than 30 different models and formulas from their research teams. An interesting term they use in their article is "contaminated data." They define it as data that includes data points from the previous year of statistics. The college basketball landscape is very fluid due to the fact that players can attend college for a single year, and then turn professional. Basketball is the only sport of the major three in North America that has this rule. In professional football, one must be at least three years removed of high school to play in the NFL. For professional baseball, they have the option to be drafted by the MLB club from high school, however if a player chooses to play

college baseball, they must wait at least three years before being eligible again for the draft or turn 21 years old before becoming eligible again for the draft; whichever of the two comes first. Because of the ever-changing landscape of college basketball, last year's data often does little to help predict this year's games. One team could make the Elite Eight one year and miss the tournament altogether in the next. This is the reason why leaving out contaminated data is so important.

The article introduces three main problems with NCAA tournament predicting. The first is that the seeds, given by the NCAA, already provide a good predictive option. The higher seeds, 1's and 2's, typically dominate their first-round games because they are playing against the lowest seeds in the tournament, 15's and 16's. Thus, historical precedent of 1 and 2 seeds indicated that they will win their first-round games. However, as the tournament progresses, the 1's and 2's will have much more difficulty winning as they play higher and higher ranked seeds. An interesting approach the authors bring up is creating an upset probability for every game and a threshold (λ), and if it crosses that certain threshold ($p>\lambda$) then an upset will occur in this game. This means that the lower ranked seed will beat the higher ranked seed.

The second problem is that there is a wide variety of methods to calculate success of models. Some methods will simply look at wins and losses of games. Others will use far more complicated methods. The third and final problem listed is that tournament historical data can have an impact on predicting success. This is tricky because historical data and precedent can often mislead predictions because of outlier years. One year, a tournament may have significantly less upsets than average. If one was using historical data to create a model, and a bracket, for this season, it would see significantly less success than in other years.

Like this project, the authors of this article used a wide variety of sources to gather their raw statistics and power rankings. Similarly, they used the statistics of Ken Pomeroy, a renowned statistician with a love for the game of basketball. Pomeroy's statistics were a major influence in this research and creation of these formulas. Pomeroy specializes in offensive and defensive efficiency and uses possessions and tempo as the basis for his statistics. Along with Pomeroy, the authors used Sonny Moore, Kenneth Massey, ESPN, and RPI, all statisticians or power ranking sites, for the newly created models of their research teams.

Of the more than 30 models, the most popular and most successful was a Logistic Regression model. The models used in the article were far more complex than what this project is doing, and interestingly, most of the models created by research teams did not pass an all 0.5 baseline test. An interesting reason they gave for why some of the models failed was that they used regular season data to predict tournament games. This was a fascinating claim by the authors because both of the formulas in this study strictly use each team's regular season data to predict the results of the NCAA tournament games.

3) The Formulas:

XPoint was the first moderately successful self-designed formula, and it was created in the fall of 2019. Since its creation, the formula has been tweaked to improve its accuracy. The formula began by only using one team's statistics, rather than multiplying Team A's statistics by Team B's to get a more accurate matchup predictor. This had some decent results, but it was obvious that it could be improved. This is when the matchup element was added to the formula. By introducing the other team's stats, it adds a more realistic element to what the game might look like. The two formulas are shown below:

(((AFTatt+BFTall/2)*AFTp)+(((ATwPatt+BTwPall)/2)*ATwPp)+(((AThPatt+BThPall)/2)*AThPp)=A Expected Score

(((BFTatt+AFTall)/2)*BFTp)+(((BTwPatt+ATwPall)/2)*BTwPp)+(((BThPatt+AThPall)/2)*BThPp)=B Expected Score

The team name (A vs B) is listed first before every variable. FT, TwP, and ThP represent free throws, two pointers, and three pointers respectively. The qualifiers “att,” “all”, and “p” represent “attempted,” “allowed,” and “percentage.” These formulas are actually much simpler than they appear in that format. First, for the three ways of scoring, free throws, 2 pointers, and three pointers, the average is taken of Team A’s attempts and how many attempts Team B allows. This gives a closer representation of how many attempts Team A will have against Team B in particular. This is important because different teams play at different speeds. For example, Virginia traditionally plays a very slow, defensive style of play, however North Carolina plays a high pace, back-and-forth style. Any given team is going to have more shot attempts against North Carolina than against Virginia. This was a major revelation in the research when this component was added to the formulas. The average shot attempts represent a likely estimate of how many attempts will occur in the game between the two teams. This average is then multiplied by that team’s shooting percentage in that category. This represents how many of the shot attempts that team will make. Finally, the amount of shots made are multiplied by how many points they are worth in the game: free throws are one, two pointers are two, and three pointers three. These three are summed to give the final expected number of points scored by the team. The same process is followed for both teams in order to predict the final score between the two teams. This is the final score that is used to determine the over/under on the game, as well as the winner of the game. In situations in which the predicted final score is between 3 points, the formula uses a self-created strength of schedule (SOS) multiplier to help determine a winner. This multiplier was only used in games in which the outcome of the game could be changed by using the SOS multiplier. The Excel table used for the SOS multiplier is shown below.

	Multiplier	Games	Total	
Gonzaga	1.09	12	13.08	
	1.06	3	3.18	
	1.03	2	2.06	
	1	12	12	
		29	30.32	1.04551724

In this table, a descending multiplier is multiplied by the number of games played by each team in that "quadrant." In college basketball, there are four quadrants for teams. Quadrant 1 is the best, and Quadrant 4 is the worst. Games played against Quadrant 1 teams represent a more difficult schedule, and thus they have the highest multiplier. The breakdown of what determines a quadrant 1 team vs a quadrant 2-4 team is shown below. The table is from Warren Nolan's RPI website, from which the number of games per quadrant per team was taken for this research.

Quadrant 1 (Q1)	**Quadrant 2 (Q2)**	**Quadrant 3 (Q3)**	**Quadrant 4 (Q4)**
Home (1-30)	Home (31-75)	Home (76-160)	Home (161-358)
Neutral (1-50)	Neutral (51-100)	Neutral (101-200)	Neutral (201-358)
Away (1-75)	Away (76-135)	Away (136-240)	Away (241-358)

These rankings are determined by RPI, ratings percentage index, which is found by adding a team's winning percentage by its opponents' winning percentage and its opponents' opponents' win percentage. The actual formula is: (0.25*WP)+(0.5*OWP)+(0.25*OOWP). The variables correspond to the phrases used above. This statistic is used in determining which teams make the NCAA tournament, as well. The total number of games played throughout the year against quadrant 1 teams is multiplied by 1.09, quadrant 2 with 1.06, and so forth as according to the Excel table.

Going back to the SOS multiplier, the number in yellow in the Excel table is multiplied by the initial final score gathered by XPoint for both teams, and then those two numbers are compared to determine a winner. Again, this is only used in cases in which the

initial final score is between three points and the outcome of the game could change due to the SOS multiplier.

XPoint focuses on the game as a whole and uses shot attempts to try to gain an insight into how many points teams will score in a matchup against any given team. Many of the earlier self-designed formulas did not focus on matchups between any two teams, but rather gave out more of a power ranking that only used that team's statistics. This had its weaknesses because it did not factor in matchups, strategy (offensive vs defensive), or specifically, tempo, which is what the next formula is based around.

KTheory was created in the fall of 2021 based on the idea that a points per possession model might take out some of the shortcomings of XPoint. KTheory is similar to XPoint in a couple ways. The first way in which the two are similar is that both give the expected final score of any game between any two teams. Secondly, it is matchup based, meaning it involves the statistics of both teams to get the predicted final score. However, unlike XPoint, KTheory focuses on breaking down the game into individual possessions, rather than looking at the game as a whole. It combines two intermediary formulas to create the final product. This formula uses more complex statistics and is not as straightforward as XPoint. The formula begins with the premise: points per possession is an important statistic to measure. The importance of points per possession is that it allows for tempo to be neutralized. Bringing back the prior example of Virginia and North Carolina, teams that play against Virginia have fewer possessions than teams who play against the fast-paced North Carolina. Points per possession takes the pace of play out of the conversation. There will not be mismatches in pace of play in this formula. The formula itself looks like this:

((APOSS+AOR)-(APOSS*BOTOpp))= AXsop

POSS=possessions; OR=Offensive Rebounds
OTOpp=Opponent turnovers per possession

(AP)/(AFGatt+(0.44*AFTatt)= AXts

Xsop=Scoring opportunities per possession
P=Points; FGatt=Field Goal attempts; PPP=Points per poss
FTatt=Free Throw attempts; Xts=True Shooting pct

Xsop*Xts= PPP → APPP*APOSS= A Expected Score

As previously said, there are a lot of variables in this formula. Beginning at the top, the first statistic that is being calculated is the scoring opportunities per possession. This represents every time a team has the opportunity to score the ball. It is much more difficult than simply using possessions. To calculate the scoring opportunities per possession, first the possessions per game is added with the offensive rebounds. Second comes the matchup portion of this formula. How many turnovers is Team B going to force Team A to commit. To calculate this, Team B's opponents' turnovers per possession is multiplied by Team A's possessions. This number is subtracted. Then the whole thing is divided by the number of Team A's possessions. This gives the amount of scoring opportunities per possession (AXsop). The second statistic that is being calculated is the True Shooting percentage. The 0.44 constant in the denominator of the true shooting formula represents the coefficient for the times players shoot 1 free throw instead of two. When a shooting foul occurs, players shoot two free throws, but there are also times in which a player shoots one free throw. These scenarios include getting fouled and making the shot or a technical foul. This number along with the number of field goal attempts is the denominator and points per game is divided by this to complete the true shooting percentage (AXts). The two intermediary statistics that were calculated: AXsop and AXts are multiplied together to get points per possession. This is the basis for the formula. With this, any two teams can be instantly compared. Finally, points per possession is multiplied by the team's possessions per game to give the final expected score. Unlike XPoint, the strength of schedule multiplier is not used in this formula for close games.

4) Method:

Now that both of the formulas have been explained in detail, this section will detail the data collection and testing process. Firstly, all statistics were taken from TeamRankings.com, Warrennolan.com, or kenpom.com. Team Rankings is a website that compiles a multitude of statistics for many different sports. It has always been accurate when compared to other sites, and it presents that statistics in a very easy to locate manner. Warren Nolan is a site for statistics and analysis for major collegiate sports. It has been active since 2001. Finally, Ken Pomeroy is the creator of kenpom.com, and he is a famous sports statistician. His website has been active since 2002. Pomeroy himself is a published author, and his work is similar to the use of sabermetrics in professional baseball. Pomeroy's statistical analysis inspired this study, and in particular, the creation of KTheory, as the formula is based on his use of tempo and possessions in his college basketball. Additionally, all odds used in this study were provided by DraftKings Sportsbook.

The data used throughout the project is both reliable and valid. The same data sources were used for each team, and the same methodology was followed for every team and every game. The formulas are definitely reliable, as they will give an expected final score every time, so long as the data input was done correctly. There is no need to retest any of the games because the same result will be reached, as long as the data input has not changed. Furthermore, the only threat to reliability would be researcher error, meaning the data was not inputted correctly. However, this was not the case, as each data entry was double checked, and there were no clear signs of error. As for the data being valid, the question that must be answered is: "How does one know that the formulas are correctly measuring what they are supposed to, and if so, are the conclusions taken from them valid?" (Gatton & Jones, 2004). The answer is yes to

both questions. The two formulas are designed to give an expected final score, albeit through very different methods. XPoint gathers a final score through expected shot attempts and then the percent of those shots that are going to be made. This gives the expected final score. KTheory breaks scoring down into scoring opportunities per possession, and then points per scoring opportunity. When multiplied together, that gives points per possession. When multiplied with possessions per game, that gives points per game. The theory is sound for both formulas, and thus all conclusions drawn from the formulas are valid. While the formulas may or may not be good indicators of predictive success, they are at the very least valid.

With the creation of the two formulas, any game between any two collegiate teams theoretically could be analyzed to predict the outcome. What was left, was to put these formulas into practice. Because both formulas were created for the NCAA Tournament, neither factors home court advantage into effect. For this same reason, they could also both be applied to most conference tournaments. In order to apply these formulas to the tournament games, first, each formula had to be put into excel. Each statistic had a specific cell, and then the cells were multiplied or divided or added as needed for the formula. XPoint is shown with the light green and the light blue, and KTheory is shown with the blue and red.

		Team B			
B FT Attempted		A FT allowed	B FT %	Expected FT Points	
B FG Attempted	A FG allowed	A 2pt allowed	B 2pt %	Expected 2pt points	
B 3pt Attempted	B 2pt attemp	A 3pt allowed	B 3pt %	Expected 3pt Points	Multiplier
				Expected B Total Points	TOTAL

	exp to per game			sop per poss	true shooting formula			TS total	Points per poss	exp points
kp adj tempo	kp adj tempo	opp to per poss	off rebounds	total 1	points	fg attempts	ft attempts	total 2	Total 2 x Total 1	PPP x kp adj tempo
kp adj tempo	kp adj tempo	opp to per poss	off rebounds	total 1	points	fg attempts	ft attempts	total 2	Total 2 x Total 1	PPP x kp adj tempo

In the NCAA tournament, 67 games are played. While the tournament has 67 games, for brackets, and the wagering analysis, there will be 63 games. This is because ESPN does not count the First Four games as a part of their bracket selection. This meant there were 67

matchups, and 68 teams worth of statistics that needed to be input into Excel. This was a massive undertaking, and to make matters more difficult, it all needed to be done in five days. Selection Sunday is the day in which the teams are selected and placed into the bracket. From Selection Sunday, the following Thursday is the first day of the tournament, meaning brackets are due and the data needed to be fully input. Additionally, data cannot be input until each team is done playing their regular season games (conference tournament games are included). The data would not be complete if it was input before the team was done playing their games. Each game a team plays strengthens their data because the sample size grows. Once the initial data is input, and the first 32 games are analyzed, then the potential games are analyzed using the predicted winners from each of the first-round games, and so forth throughout the rest of the tournament. From this, two brackets were created: one from the predicted outcomes of every game using XPoint, and the other from the predicted outcomes of every game using KTheory. The limitation of the bracket is that it cannot be updated. After the first-round games, there are potentially different matchups than were predicted by the formulas, and thus that game was not analyzed in the original predictions. However, for the sake of a larger sample size from which to analyze the formulas, the actual matchups that occurred in every round were also analyzed. If only the games predicted in the bracket, many of these games would have never been analyzed. This way, the sample size truly is 63 because every game played was analyzed whether the formulas predicted the matchup or not.

In addition to this, each game of the tournament was analyzed from a sports wagering perspective. Each game was analyzed in three ways: on the money line, against the spread, and the points total. On the money line simply means picking the winner of the game. Against the spread means the bet is occurring on a predetermined point spread instead of simply betting on

the winner of the game. The point spread is determined by sportsbooks with the intention of having 50% of the money wagered on a game be on each team. If this occurs, the sportsbook wins no matter the result of the game due to the "vig," which is the percentage the sportsbook keeps on winning bets (usually 10%). More on this method of wagering to come later. Finally, the points total represents the combined final score of the game. Then bettors will bet over the sportsbooks' total, or under. Through the lens of each of these three methods of sports wagering, each game of the tournament was analyzed using both XPoint and KTheory. Additionally, to accurately portray how the formulas would do in the sports wagering realm, imaginary $100 bets were placed on every game and tracked for each of the three methods and two formulas.

5) Results:

Beginning with the brackets because they were the primary inspiration for the study, the brackets were submitted on ESPN's Tournament Challenge. Inputting the stats for every game into the formulas on Excel gave the predicted final score of the game, and thus, the winner of the game. This determined which team was chosen to win in the bracket. For first round games, XPoint correctly predicted 17 out of the 32 games. KTheory correctly predicted 19 out of the 32 games. For first round games, this is not overly successful. There is not a 50% probability of each team winning the game, due to the seeding of the teams in the tournament. Number one seeds have only lost one single game in the history of the NCAA tournament in the first round. Two seeds have only lost ten total times in the first round of the NCAA Tournament (Wilco). Therefore, one and two seeds have a much higher than 50% chance to win their first-round games, making the 17 and 19 first round wins for the two formulas much less impressive. Additionally, with the 15 and 13 losses, the quest to create the perfect bracket was extinguished very quickly.

The second round is where analyzing the brackets themselves gets a little tricky. Because the formulas had so many losses in the first round, including a team predicted to go to the Championship game in XPoint, there were second-round matchups that were not analyzed with the formulas. An example of this is Gonzaga vs Memphis. In the first round, Memphis played Boise State. XPoint incorrectly predicted Boise State to win the game. In the bracket, Gonzaga was predicted to beat Boise State. Gonzaga won the game against Memphis, and thus the bracket was awarded points because Gonzaga was correctly chosen to advance to the Sweet 16, even though the bracket had them beating the wrong team. So, rather than focusing on the amount of matchups each formula won or lost, when analyzing the bracket, it is more important to look at how many of the 16 teams that advanced to the next round were correctly predicted. XPoint correctly predicted 7 of the 16 teams that advanced to the Sweet 16. KTheory did slightly better, correctly predicting 8 of the 16 teams. Of the 16 teams, 10 of them were "chalk," meaning the higher seed (lower in number, higher in "ability") won both games they played to make the Sweet 16. XPoint correctly predicted 2 of the non-chalk teams to make the Sweet 16. While, this is actually quite impressive, the problem was that XPoint missed 5 of the chalk teams that reached the Sweet 16. This can be generalized into the larger issue of XPoint predicted too many upsets. A possible issue of this was that the SOS multiplier was only added into effect at the end of the formula rather than integrated fully into the calculation of the formula. However, this could skew the expected final score, thus having an effect on correctly wagering on the points total of the game.

Moving onto the Elite Eight, both XPoint and KTheory only correctly predicted one team to make the Elite Eight. This was disappointing to say the least. While this was a tournament filled with upsets, and not one Elite Eight matchup was chalk, it was still

disappointing to see the lack of correct picks to get to this stage of the tournament. Past the Elite Eight, neither XPoint nor KTheory predicted any of the Final Four teams, and thus neither of the teams that made the Championship game.

The two brackets ended with 350 and 390 points. The points structure in ESPN's Tournament Challenge awards 10 points for first round wins, 20 for second round wins, 40 for Sweet 16 wins, 80 for Elite Eight wins, 160 for Final Four wins, and 320 for correctly picking the National Champion. There is a maximum of 320 points each round, and there are 6 rounds for a maximum total of points of 1920. Compared to other players in ESPN's Tournament Challenge, XPoint came in the 9th percentile of brackets, while KTheory did slightly better coming in the 16th percentile of brackets. These are dismal results. As far as the brackets performed, it is safe to say they were a failure. However, there are many other ways in which the success of the formulas can be measured.

As said previously, in order to ensure a large sample size, before every round, the games were put into Excel and the results were predicted. The original predictions had no effect on these games because each round was its own separate analysis. Therefore, each of the formulas analyzed the 63 games of the tournament. Strictly trying to pick the winner of the games, XPoint had a record of 36 wins and 27 losses, while KTheory came in slightly better at 38 wins and 25 losses. These win percentages are not bad, around 57% and 60% respectively, but once again, because most of the games have an uneven win probability, the idea was to improve above a 60%-win percentage. In scenarios of 50-50 games, 60% accuracy would actually be fairly strong. This leads into the sports wagering section of the study. How did each of the formulas do when applied to the three main methods of sports wagering?

6) Sports Wagering:

a) Money Line

Betting on the money line is simply picking the winner of the game. This is popular in sports without high margins of victory such as baseball, hockey, and soccer. In sports like basketball and football, where the points spread is bigger, betting against the spread is more common and more profitable, as shown later. Doing the money line calculation was easy to get the record. It was simply comparing which team the formulas had predicted was going to win the game with the team that actually won the game. This gave the record of 36-27 for XPoint and 38-25 for KTheory. However, this might lead one to believe that the formulas were profitable against the money line. This is simply not the case. When betting on the money line there is a favorite and an underdog. The favorite will have their odds listed at "-120" for example. This represents the amount of money needed to be wagered in order to win $100. So, in this case, one would need to wager $120 in order to win $100. Because they are betting on the favorite to win the game, naturally they are not going to have as high of a payout. For the underdog, their odds will look like this: "+150." This is the amount of money that the bettor would win on a $100 bet. Because the bettor is choosing the underdog, the team that is less likely to win, the payout is much larger than choosing the favorite. This is the exact reason why betting on the money line is not profitable in March Madness. The one and two seeds are extremely heavy favorites, and on a $100 bet, they offer single digit dollars in return if the bet wins. Whereas, if the bet loses, the bettor loses $100. There is no value in these types of bets unless betting on the underdog. In this year's tournament, both Baylor and Gonzaga were -10000 to win their first game. They both won, and thus on two imaginary $100 bets, the profit is $2. There are not enough wins by the favorites to cancel out the inevitable losses. Also, the tournament is called March Madness for a reason. The biggest underdog to win a game in this year's tournament was Saint Peter's against

Kentucky. Kentucky was -4000 to win the game as one of the 2 seeds. Saint Peter's was +1400 meaning on a $100 bet, the bettor takes home $1,400. Because of the complete lack of value in money line betting in the NCAA Tournament, when using a standard unit bet for every game, this is not an appropriate way to determine the success of the formulas.

b) Against the Spread:

Betting against the spread is the most common method of gambling in basketball and football. This is because the margins of victory can have a significantly larger spread than those of baseball, hockey, and soccer. A blowout in those sports would be winning by five, however winning by 5 indicates a close game in both basketball and football. So, what does betting against the spread mean? The spread represents the margin of victory predicted by the sportsbooks. However, when sportsbooks create a spread, they are putting out a projected margin of points between the two teams that will cause the public to bet 50% of all money on both teams. This ensures the sportsbook makes a profit because the sportsbook takes 10% of the money for accepting the bet. Therefore, betting against the spread is simply betting against the sportsbooks' projected margin of victory. What is different about betting against the spread is that the underdog does not have to win the game to cover the spread. If Team A is favored by 6.5 points (the half point ensures that no bets tie), then Team B is the underdog by 6.5 points. If Team B loses by 3 points, they lose the game, but cover the spread, and therefore, the bettors that bet on Team B to cover the spread win their bets. Essentially all against the spread bets have the odds of -110 for both the favorite and the underdog. All imaginary bets in this study were placed with the implied -110 odds against the spread.

The method used to pick winners against the spread was as follows. Because the formulas are designed to give expected final scores, those can be compared to the actual spread,

as given by the sportsbook. This can get tricky when the predicted winner of the game (Team A) was predicted to win by less than the sportsbooks' spread. In this case, Team B was predicted to cover the spread, because in the formula, Team A did not cover the sportsbooks' spread. One example of this scenario occurring is the first-round game between Baylor and Norfolk State. Both formulas predicted Baylor would win by single digits, 6 points in XPoint and 7 points in KTheory, but the spread was 20.5 points. Therefore, because in the predicted outcomes, Baylor did not cover the spread, the "bet" placed was on Norfolk State to cover the spread. This is one scenario in which choosing this method worked against the formula, as Baylor won the game by 36 points. One game in which using this method was favorable, was Kentucky against Saint Peter's. XPoint and KTheory predicted that Kentucky would win by 9 points and 6 points respectively. However, the spread was 18.5 points. Therefore, the bets placed were on Saint Peter's to cover the spread, and they won the game outright. This method is consistent for both formulas' results.

Up to this point, KTheory had outperformed XPoint in its accuracy of predicting games on the money line. However, surprisingly, XPoint was far superior in predicting games against the spread. XPoint's record against the spread was 36-27, for a win percentage of approximately 57%. This is a solid win percentage against the spread. In order for a model to be profitable against the spread, it must win at least 55% of games. It is higher than 51% because the sportsbook takes 10% of winning bets. Therefore, winning a $100 bet gives a payout of $90.91. On the 63 games of the tournament, against the spread, with imaginary $100 bets on every game, XPoint won $572.73. This was a very promising result to see after the disappointing of the brackets and the money line results. XPoint got off to a rough start, going 16-16 in the first-round games, but recovered very well. After the first round, XPoint correctly picked the

winner against the spread in 20 of the following 31 games. Unfortunately, KTheory did not perform as well. Going 28-35 against the spread was definitely surprising after its relative success on the money line. KTheory recorded a loss of $954.54 on its bets against the spread. Another disappointing result for the two formulas. However, XPoint's success marked a high point in the research so far.

c) Points Total

The points total method of gambling is quite popular as well. Fairly easy to calculate, the points total is the combined score of both teams. This is a pretty universal method of gambling because the total can be so easily changed depending on the sport. Totals in baseball range from 6.5 to about as high as 10.5 total runs. NBA games routinely have totals in the low 200s. For the games of this NCAA tournament, the totals ranged from 123.5 to 156.5. The results for the points total from the two formulas were slightly unexpected. Starting with XPoint, its record on the points total was 29-33-1 with a loss of $663.64 on its bets. Shockingly, XPoint predicted higher than the points total on every single game played in the tournament. This was completely unexpected and unprecedented, as it meant XPoint was predicting every game to go over the total. It just so happened that the more games went under this year than over. Placing bets on the under on every game would have gone 33-29-1. The one game pushed, meaning the combined score of both teams landed directly on the total line given. The bet is cancelled in this scenario. Because XPoint predicted the over in every game, it served as a sort of control from which to compare KTheory's results.

KTheory outperformed XPoint again in the totals, after it fared much worse against the spread. It went 33-29-1 with a payout of $100 on the dot. Earning a profit was impressive, but it still was a bit too skewed towards predicting the games to go over. Of the 63 games, it

predicted only 10 to go under. Of those 10 games, seven of them went under. Going 70% on those bets is very impressive albeit is with a small sample size. Even with the tendency to predict the over, KTheory still turned a profit on the totals, so this is chalked up as a minor success.

d) Underdog Winners ATS

When analyzing the data, a question arose: "Were the underdogs that the formulas chose to win, at least competitive in their games?" The answer to this question would show if the formulas were fundamentally flawed, or simply need tweaking. The way this question was answered was by looking at every matchup in which the formulas predicted a lower seed to upset a higher seed. In these games, it was important to know if the lower seeds that were predicted to win the game, were at least covering the spread. There were 17 of these games in XPoint, and 14 in KTheory. Each formula had a winning record: 10-7 for XPoint and 8-6 for KTheory. Both formulas recorded winnings on their bets, $209.09 for XPoint and $127.27 for KTheory. These results show that the formulas were not inherently flawed as the teams they predicted to make upsets, at least covered the majority of the time. Furthermore, this could be a profitable model to follow in the future considering both formulas ended the tournament with a profit rather than a loss. Unfortunately, there are not a lot of games that fall into this category as it was only around a quarter of all the tournament games.

7) Applications:

What can these formulas be applied to? To begin with, as shown, each formula has its merit in some form of sports wagering, XPoint with against the spread bets and KTheory with totals. But outside of the world of sports wagering, can this be applied to anything else? An argument can be made that coaches can use this to create more efficient teams.

XPoint is based off of what percentage of shots is a team going to make of the number of attempts the other team allows. Coaches can use this theory of basketball to try and limit three pointers against a strong three-point shooting team or try to limit free throws against a team that gets to the free throw line often. Or conversely, knowing that a team does not shoot three pointers at a high percentage, the coach could create a gameplan that allows the other team to attempt more three-pointers, but limit the amount of two-pointers and free throws. Some coaches have even abandoned playing defense at all. In a Division III NCAA basketball on December 8th, 2022, the Grinnell College Pioneers attempted 111 three-pointers in a 124-67 win. Every shot the Pioneers took was a three-pointer, aside from the four made free throws of six attempts. They made 40 of the 111 attempts for a three-point percentage of 36% from beyond the arc. Grinnell plays a strategy in which the shoot quick three-pointers, and everyone besides the shooter attempts to get an offensive rebound. On misses, they give up an easy layup in order to get the ball back quicker. On makes, they use a full court press in order to try and force turnovers. If the other team breaks the press, Grinnell allows the easy layup and goes back to taking quick three-pointers. The theory is that they will make enough three pointers and force enough turnovers to make up for the easy layups they give up (Associated Press). This coach understands that the more three pointers his team attempts, the more possible points the team can get. After that, it is up to the team to make a decent percentage of them.

KTheory can be applied by coaches who are looking to maximize their efficiency. KTheory focuses on points per possession. Teams with a high points per possession statistic can play at any pace, and still score efficiently. Coaches should try to increase their team's points per possession by preaching quality shots instead of a high quantity of shots, in this case the opposite of what Grinnell College does. Teams with a high points per possession can play at the slow,

methodical, defensive pace, or they can play at a high octane, fast-paced, offensive pace. More coaches should emphasize this in their gameplan and strategy.

8) Future Research and Shortcomings:

This year's attempt to find the perfect formula to maximize the success of March Madness brackets provided some very good insight. This year's research has certainly been more extensive than any of the other years and it has absolutely served as motivation to continue the process. In next year's research, there will be some sort of game theoretic model added, based upon the tempo at which teams play. This should help differentiate the two distinct play styles, and hopefully predict how teams will respond to certain matchups.

As for the shortcomings of this study, there is a slight feeling of disappointment with the results from this tournament. Every year, the formulas enter the tournament was so much promise and hope, however this past year had felt especially hopeful. The number one thing that will be assessed in the offseason before next year's research is the use of effective field goal percentage instead of the true shooting percentage in KTheory. It's possible that effective field goal percentage might more accurately portray the 'points per shot' part of the formula that was trying to be achieved. Additionally, there will almost definitely have to be some tweaks to XPoint's application to totals. My hypothesis on why XPoint predicted every game to go over is that teams are not as defensively concerned in the regular season as they are in the NCAA tournament. When teams are fighting for their season, they are going to be more inclined to play harder defense. This will definitely need to be investigated further in next year's research. Finally, XPoint and KTheory both need to find a way to accurately integrate strength of schedule into their formulas. Norfolk State and Baylor do not play with the same schedule difficulty, and

thus Norfolk State's statistics in the regular season are going to be better than they would be if they played Baylor's schedule.

9) Conclusion:

In conclusion, although XPoint and KTheory came into this year's NCAA Basketball Tournament providing high hopes, the ever-elusive perfect bracket was not achieved. However, both formulas had their merits in certain aspects. XPoint performed very well against the spread, while KTheory performed above average with totals. Overall, both formulas need to be tweaked in order to try to be more accurate. Whether it is possible to create a perfect bracket or not, the quest will continue year after year, and maybe one day, it will not end in disappointment. That sounds much more depressing than it actually is. March Madness is a favorite sporting event to watch for many, and every year people will sit down and watch as the madness unfolds and count down how long it takes for their bracket to end up in the paper shredder.

References

Gratton, C., & Jones, I. (2004). Theories, Concepts and Variables in Sport Research. In *Research methods for sport studies* (pp. 71–90). essay, Routledge.

Karnibad, Jarrett. "March Madness 2021-2022." https://docs.google.com/spreadsheets/d/1I3OOlJA3XfgqcrVu9cTTag0xz3nEyLisUNvlpFxbedM/edit#gid=293515309

Press, Associated. "Grinnell Sets NCAA Record with 111 3-Point Shots vs Emmaus." *The Washington Post*, WP Company, 9 Dec. 2022, https://www.washingtonpost.com/sports/colleges/grinnell-sets-ncaa-record-with-111-3-point-shots-vs-emmaus/2022/12/09/44ae4332-77e0-11ed-a199-927b334b939f_story.html.

Wilco, Daniel. "How Men's Basketball Teams That Earn a No. 2 Seed Do in March Madness." *NCAA.com*, NCAA.com, 13 Mar. 2022, https://www.ncaa.com/news/basketball-men/bracketiq/2021-03-10/march-madness-getting-know-no-2-seed-ncaa-tournament.