GOATs and BOATs; or When Might 11/13 be Less Than 6/18?

Rick Cleary: Babson College: rcleary@babson.edu
Steven J Miller: Williams College: sjm1@williams.edu

Abstract: Comparing extraordinary results and determining GOATs (Greatest of All Time) is a favorite discussion topic for sports fans and data analysts alike. The analysts, however, are more likely to recognize that attempts to measure greatness are always sensitive to the metric chosen. We consider the general questions that arise when those metrics are associated with the "unlikeliness", or right tail probability, of team success and how much of that success is associated with a single player. We would love to eventually answer, or at least offer a thoroughly supported opinion, on who is the GOAT of all GOATs across team sports. This paper spells out preliminary steps in how this could be done. Because it is difficult to compare players of different positions in the same sport, and because we must also compare across not only sports but eras, we narrow our focus to finding the BOAT of all BOATs (Best of All Teammates). We are interested in who has *seen* the most success, not necessarily who has contributed the most. We outline a framework using a specific example comparing Tom Brady and Bill Russell, two popular candidates for BOAT (or GOAT), and discuss in detail the challenges of comparing across sports, in particular how the playoff structure affects the metrics, and we come down, narrowly, on the side of Brady. We encourage readers, especially students, to join us in considering specific cases and extensions of these ideas and we hope this discussion will be a fruitful source of projects for classes or independent studies in probability, sports analytics, or modeling.

Section 1.) Introduction

New England sports fans of the baby boom generation have been very fortunate to enjoy two long periods of dominance by their local teams. From 1957 through 1969 the Boston Celtics won eleven of the thirteen National Basketball Association (NBA) championships. (For basketball, we consider the "1969 season" to mean the season in which the playoffs were held in 1969). Much of this success was attributed to their center Bill Russell, who was also the coach of the team during the tenth and eleventh championships. A few decades later, the New England Patriots entered a period of sustained excellence, winning six National Football League (NFL) championships in eighteen seasons from 2001-2018. (Since most of the football season is played in the fall, we use 2001 season to mean the regular season was in 2001, the Super Bowl was in 2002. We apologize for the lack of consistency with the NBA, but this appears to be the convention.) Quarterback Tom Brady was the acknowledged star of these teams. These successes plus the seven combined championships by our baseball team (four) and hockey team (three) has been almost too much good news for taciturn New Englanders to bear!

At first glance, it seems clear that the Russell/Celtics streak is more impressive, but in this paper we present some background about the structure of these leagues and their playoff format that suggest that the conclusion is not obvious. We use probability models to study the relative dominance of teams in their league, and in future work we hope to consider the contribution of individual players to the team's success. Using the careers of Russell and Brady as our primary examples, we wish to develop a widely applicable approach to evaluate which players across sports were associated with the greatest amount of team success. While acknowledging that no such evaluation system will be perfect, and all will be sensitive to choices made, we believe that measures developed using sound statistical and probabilistic modeling can shed some light on the relative success of teams from different sports and eras.

Since our focus is on team success, we are coining the term BOAT (Best of All Teammates) as a slightly more quantifiable alternative to the GOAT (Greatest of All Time). In addition to shedding light on a sports question that people care about, our approach also provides a good teaching tool for courses in probability or sports analytics, as there is ample opportunity to refine and generalize. Throughout this paper we demonstrate our ideas using Russell and Brady, who are routinely debated but not always with quantitative arguments. (See [1] for an argument for Brady, and [2] for an argument for Russell.)

A note about future work on the impact of an individual: Despite the development of 'contribution statistics' like WAR (Wins Above Replacement), there are innumerable challenges in such an analysis, as in team sports an individual's statistics are greatly influenced by their teammates. For example, if a quarterback has a weak offensive line, no credible running game, or is constantly starting deep in his own territory, this will greatly impact their statistics. The difficulties grow when we compare different sports and different eras; how to account for the effect of free agency, salary caps, and playoff formats, to name just a few. We conclude with a list of open questions (Section 5) that includes suggestions for how to address the 'individual contributions' problem.

Before moving on, we must first define our metric for what counts as success? In some markets it is a zero-one scale: did we win the championship, or did we fail to do so? With such Draconian metrics four straight Superbowl losses (apologies to any Buffalo Bills fan reading this) would rate the same as finishing last each year, and much less than one ring and three winless seasons.

For this analysis we present preliminary results using league championships as our primary measure; a natural project is to extend and give additional points the further one advances in the playoffs (as well as season milestones, such as being undefeated or having the most points ever), opening up the fascinating question of what weights to use, and how stable our rankings are relative to these choices.

Section 2.) Celtics and Patriots: A team-based model with league factors

At first glance, it appears that the Celtics' accomplishments with Russell are far more impressive than the Patriots with Brady. Winning eleven titles in thirteen years seems to dominate six titles in eighteen (or if we consider his years with Tampa, seven in twenty-one). However, this naive approach neglects two key factors, both of which bolster the case for Brady and the Patriots. We build these into probability models as follows.

Factor 1: Number of competitors. The Celtics played in a league that had as few as nine teams, usually ten, while the Patriots always competed in a league with thirty to thirty-two teams. We build a simple first model based on this information.

Model 1: We assume each season begins with each of the n teams in the league equally likely to win the championship. Letting X be the number of championships won in k seasons, P(X = x) follows a binomial(n, 1/k) model. For our example:

- the probability the Celtics win at least 11 of 13 in a ten-team league is $6.4 * 10^{-10}$, and

- the probability the Patriots win at least 6 of 18 in a 32 team league is $1.25 * 10^{-5}$.

While the Celtics still seem to have done something much less likely, we have to consider another important difference between the NBA and NFL.

Factor 2: Playoff series length. NBA playoff rounds in the Celtics era of dominance consisted of a series of games, usually best of seven but sometimes best of five. The NFL playoffs, however, are a single elimination tournament. Playing a series rather than a single game reduces the probability of an upset and a superior team being eliminated, though perhaps not to the extent that many fans believe. (This is detailed here in section four, see also [3].)

Reviewing the Celtics championship run, we see that there were four times during this window that they lost the first game of a playoff series that they went on to win. What if those series had been 'single elimination' instead of a series? Then the Celtics have just seven titles in 13 years in a (usually) 10 team league. Some might argue that the Celtics would not have lost those 'game ones' if they know they were elimination games, perhaps by playing starters more minutes, but the historical record does not suggest the Celtics ever changed strategy or were resting a regular in those games that were well before the era of load management.

Model 2: We penalize the Celtics for losing the first game of a series. Now the probability the Celtics win at least 7 of 13 in a ten-team league is $9.95 * 10^{-5}$. That's the same order of magnitude as the Patriots, but not quite as unlikely!

With just two simple and reasonable considerations, we have 'leveled the odds' for the Celtics and the Pats.

Two other basketball teams deserve a mention before we look at other sports. First, the 1990's "Michael Jordan Chicago Bulls" won six NBA championships in just eight years. If one discounts the one season Jordan missed entirely and another where he returned very late in the regular season, one could consider them a "six wins in six years" team at a time the league had 27 to 29 teams. However, they had first game losses in three of this six championship runs, so our second factor above would have sent them home early.

The other basketball team that deserves consideration is the Coach John Wooden UCLA dynasty that produced eleven NCAA Division I championships in thirteen seasons, just like the Celtics! A future line of research of how to compare college and professional sports teams is suggested in our questions in section 5.

Section 3.) Other candidates

Our study of this issues was motivated by teams with many championships in a relatively short time period, especially those teams associated with an individual like the "Russell Celtics" and the "Brady Patriots." But are those really the best examples? We think that perhaps they are the best, as a review of Major League Baseball (MLB) World Series champions and National Hockey League (NHL) Stanley Cup winners does not reveal a team whose dominance over a period of time was quite as impressive as the Celtics and Patriots. Likewise, none of the team sport candidates suggested at [4] seemed to hold up upon inspection.

In the case of MLB, the New York Yankees run of six championships in seven years from 1947-53 is the top candidate. These teams are not strongly associated with one superstar but with a pair of great outfielders, as Joe Dimaggio retired after the 1951 season, just before Mickey Mantle burst onto the scene. (Yogi Berra might now be considered the star of those teams, though it wasn't until the mid 1950's that he became thought of that way, and after 1953 the Yankees won just twice in the next seven seasons.) The fact that this was an era with no interleague play raises an interesting modeling question: Using the assumption of teams being equally likely to win the pennant, and seasons being independent of each other, should we treat at least six titles in seven years as a two stage process, where we first compute the probability of winning an eight team league and then assign a probability of one-half to beating the National League team in the World Series? Or should we just compute the probability of a team winning a sixteen-team league at least six times in seven years?

Fortunately, this choice makes almost no difference! Under the two-stage probability model, we find the probability of at least six titles in seven years to be about $3.75 * 10^{-7}$. The "one sixteen team league" approach yields $3.95 * 10^{-7}$. Is that exponent of negative seven enough to say the Yankees accomplished something more unusual than the Patriots or Celtics? Not yet, because we haven't accounted for playoff series length; which was 'factor 2' in our previous section. During two of their championship seasons, the Yankees lost game one of the World Series. Once we make this adjustment, we find the probability for the Yankees grows to $4.58 * 10^{-4}$. That's an impressively rare feat, but not as impressive as our section two contenders.

The situation in hockey is slightly different. Because the NHL had only six teams for many years, fewer teams than the NBA, a string of championships would have had to have been even more impressive than the "11 of 13" put by the Celtics to be a contender. No such dynasty emerged. The league doubled in size to 12 teams by 1967, and by 1980 had grown to 21 teams. At that time, two powerhouse teams rattled off impressive eras of dominance.

The New York Islanders were the first, winning four straight Stanley Cups from 1980 to 1983. Using our simplest model we would estimate this probability simply as $(1/21)^4$, an impressively small 5.14 * 10^{-6}. But the pesky 'series length' factor would have eliminated one of the cups, as they lost the first game of a playoff round. Including this changes their win probability to 4.16 * 10^{-4}, very similar to the DiMaggio/Mantle Yankees. Impressive, but not quite a contender.

Immediately following the Islanders, the Edmonton Oilers of Wayne Gretzky won five Stanley Cups in seven years, all in a 21-team league. This event also has a "10^{-6}" order of probability, but the Oilers would not have survived a single elimination tournament in three of their five winning seasons, and thus do not appear to be a contender using our methods.

Each of the multi-time champions mentioned so far did something very impressive. An important 'across sports' interpretation of our results is that the very small probabilities we assign to each team's outcomes is based on the model that assumes equal likelihood for each team winning in a particular year. This assumption is clearly false as some organizations are simply better than others at acquiring and developing talent. The probabilities we compute are effective for comparison, but they are not meant to be used for prediction of future championship runs.

Section 4.) Additional Detail on Series Length

The discussion above shows that we need a good way to compare the effect the number and lengths of series have on winning a championship. As remarked earlier, the longer the series the greater the chance that the better team triumphs and avoids an upset. We wish to quantify this obvious observation to get conversion factor between titles in various leagues. For example, how many Superbowl rings would 11 NBA titles equal?

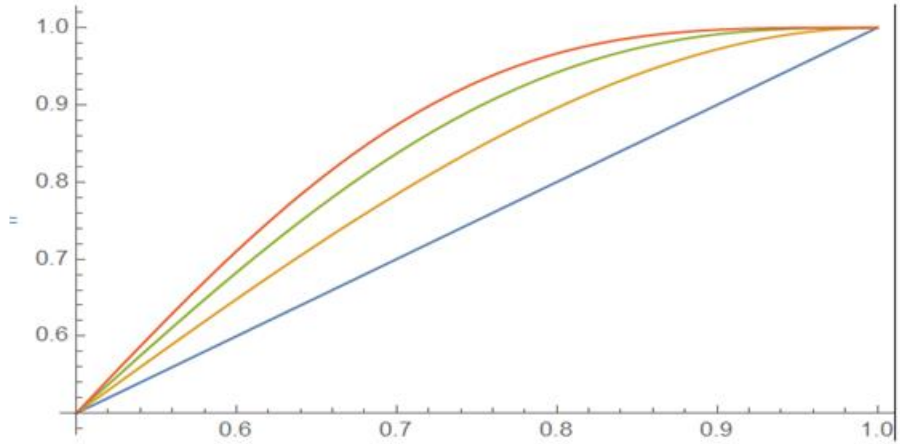
4.1) Series Length Preliminaries

Of course, the answer will depend on both the number of teams in the leagues as well as the number and lengths of each series. Thus we start with a simpler question: if a team has a probability p of winning each game, and each game is independent of the others, what is their probability of winning a best of 2n+1 series? In other words, what are the odds that they win n+1 games before the other team wins n+1?

The answer is a sum of weighted binomials. For our team to win in exactly m games, they must win n-1 of the first m-1 games, and then win the m^{th} game. Therefore they win the series with probability

$$\sum_{m=n+1}^{2n+1} \binom{m-1}{n-1} p^{n+1}(1-p)^{m-(n+1)}.$$

We plot this probability below; the x-axis is the probability the stronger team wins a game and the y-axis is the probability it wins a series, with the blue curve being a best of 1, the orange a best of 3, the green a best of 5 and the rest a best of 7. As expected, the best of 1 is just a line, the longer the series the greater the chance of the better team winning, and for all curves the probabilities are 50% and 100%, respectively, in the extreme cases when p = ½ or p = 1.



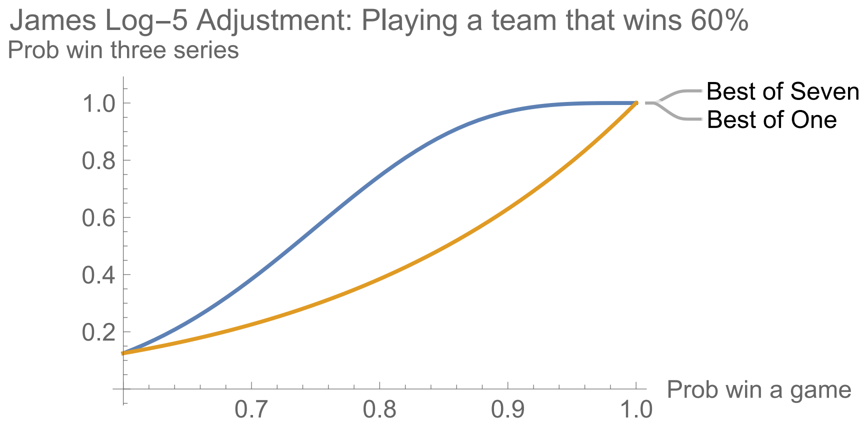
4.2.) An alternate approach using the log-5 method

Now that we can compute the probability of a team winning a series of a given length when we know their probability of winning a game, next we must determine what is the probability of winning a game. For simplicity we assume our team is equally likely to win all games, thus ignoring the effect of home field / court advantage. We use Bill James' log-5 method: if team A wins p_a percent of their games and team B wins p_b, then the probability that A beats B is

$$Prob(A\ beats\ B) = \frac{p_a(1 - p_b)}{p_a(1 - p_b) + (1 - p_a)p_b}.$$

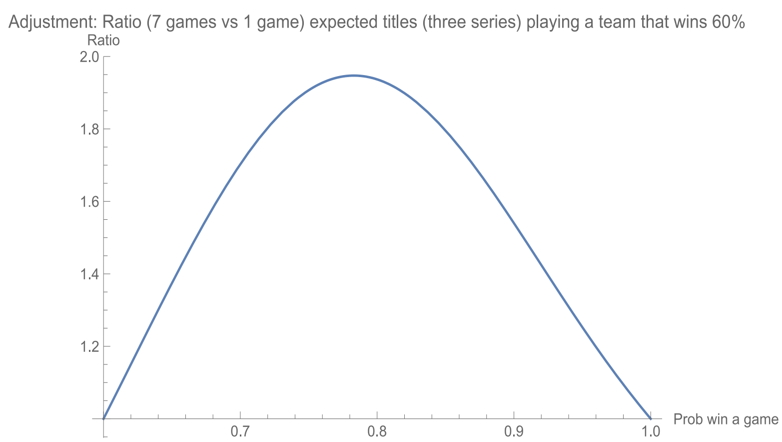
For example, in a playoff round we might have two teams where team A has won 80% of their regular season games and B has won 60% of theirs. (Most of the time each playoff team should have won more than half their regular season games, though in leagues where a high percentage of teams make the postseason this is not always the case.) Substituting into the equation above yields the probability of A beating B in a game to be approximately 72.7%. This is a reasonable result; it exceeds 50% (as it should since A is the better team, with a greater winning percentage), but it is less than A's regular season win percentage of 80%, indicating that A is now playing a better than average team.

We combine these two steps to determine the probability a team wins three series. For definiteness we assume Team A's opponents always win 60% of their games, and compute for A winning from 60% to 100% of their games with each series either a best of one or a best of seven. Note if A wins 60% of their games we expect them to win each series with probability ½, and thus the probability of winning three series would be $(1/2)^3$ or 12.5%.



Not surprisingly, as Team A's probability of winning approaches 50% or 100% the series length does not matter, but for probabilities around 80% the difference is quite pronounced.

To quantify this difference, let's look at the ratio of how many times Team A wins three best of seven series versus three best of one series as a function of their probability of winning a given game.



When their probability of winning a game is around 80%, we get a ratio of about 1.9. We can thus use this to try to convert from NBA to Superbowl titles. Assuming three series for each, with a best of seven each time for basketball versus one game do or die in football, and assuming the better team wins around 80% of their regular season games and plays opponents winning 60%, then for every 1.9 NBA titles we expect 1 Superbowl Ring, so 11 NBA titles would equate to a little less than 6 NFL championships.

It is important to note that this final conversion is the result of numerous simplifying assumptions; we are ignoring home field, and we are assuming the strength of the opponent is independent of the round. A more involved analysis can take into account the fact that it is more likely to have better opponents deeper in the playoffs, but as our goal here is to just get a rough sense, we content ourselves to this simple model.

In particular, the 11 rings of the Russell era no longer look overwhelmingly better than the 6 earned by the Patriots (though it was done in fewer years).

A standard conditional probability argument gives a good justification for James' log-5 rule, which suffices for our simple analysis (one could of course use more advanced statistics, such as his Pythagorean expectation; we will explore that in a later paper). We model teams A and B playing as follows. With probability p_a team A has a good performance and with probability $1-p_a$ a bad one, with the probabilities for B having a good or bad one being p_b and $1-p_b$. We have A and B keep playing until one does well and the other poorly, and declare whomever played well the winner. We can model this as a memoryless process, as if both have a good or both have a bad performance then it is the same as that event did not happen. Note the probability A beats B is reduced to a simple conditional probability (we are conditioning on the two teams do not both play well or both play poorly). The probability A plays well and B plays poorly is just $p_a(1 - p_b)$, while the probability both have different outcomes is $p_a(1-p_b) + (1-p_a)p_b$; the ratio is the desired estimate for the probability team A defeats team B.

The details in this section show that we can take a rigorous approach to studying the probability of sports outcomes. The general procedure for this sort of work is standard in all probability modeling. First, carefully define the problem, and begin with simple assumptions. Then, experiment with relaxing assumptions and checking if it makes a significant difference in the results. If relaxing assumptions lead to the problem not being solvable in closed form, consider simulation. Then, when possible, compare the model results to any available empirical data. The next suggestion gives alternatives for some follow-up questions that could use this approach.

Section 5.) Suggested Questions

The search for GOATS and/or BOATS is a wonderful source of interesting questions for classes, honors projects, independent study projects, sports analytics clubs, and of course pubs and parks. While we propose the idea of comparing success through probability modeling, there may be other metrics that could be illuminating. A preliminary list of questions follows.

1.) Our models in this paper are based on counting championships. How would the results look different if we gave "partial credit" for playoff success in preliminary rounds, and how sensitive are these results to the weightings assigned the relative importance of each round?

2.) Are our examples of the 1957-69 Celtics in the NBA and the 2001-2018 Patriots in the NFL the right choice for the teams to consider? Are there assumptions about time period or league size that would suggest other candidates?

3.) What teams are candidates in other sports and other levels and what individuals are associated with those teams? Consider for example NCAA teams, women's college or professional sports, or Olympic level sports. (This is a great question for encouraging research into women's results, which are too often overlooked.) How does the structure of these other events compare to the professional playoff structure we consider in sections two and three?

4.) How might we measure and account for the level of roster stability? Clearly this is related to the assumption of independence used in the simple models we have looked at so far. If teams return for a new season with a roster almost unchanged, this is very different than having a lot of player movement. The extent of that movement has changed considerably over time. A good example in the case of the Celtics is that Bill Russell's teammate Sam Jones, also a Hall of Fame caliber player, was present for the last ten of Russell's eleven championships; while in the Brady era players had considerable freedom and moved more quickly.

Related to individuals, we consider the following:

5.) Sport Structure: When Bill Russell played, he was in the game for the Celtics about 80% of the time, and he was one of five players. Tom Brady was on the field only when the Patriots had the ball, and even then he was just one of eleven. On the other hand Tom Brady played quarterback, widely acknowledged as the most important position on a football team, and one in which he touches the ball on every play. Measures like WAR (Wins Above Replacement) have been developed to estimate an individual's contribution to team success and might be included here. Good resources on individual contributions are available for basketball ([6] and [7]) and the much more complicated issues that arise in football [8]. The football problem is complicated by the vast difference in how much measurable data is available for different positions. The "skill positions" of players who handle the ball have abundant data while linemen have little individual data.

6.) If a BOAT candidate moves and wins elsewhere, should that be weighted in their favor? Just as an example, Bill Russell stayed in one place for his entire career, while Tom Brady and LeBron James moved and won again; twice in James' case. On the other hand, some stars (Michael Jordan, Wayne Gretzky) moved and did not win.

7.) How much credit should be given to coaches/managers as well as to players? Could our models extend to them as well? In particular, Yogi Berra had success both as a player and as a manager.

8.) There are other team factors that could be considered. One important one is roster stability. The Celtics titles came in an era when player mobility was extremely limited and so they maintained the same core group for a long period. Their opponents did likewise. The Patriots' success came after the advent of free agency when personnel changed much more quickly. How to account for this in probability models is one of our key follow up questions.

9.) Who does the public think is the greatest of all time? For those who believe in wisdom of the crowds, perhaps polling is the way to go. One can see an example of this approach in [5], but for students it would be interesting to form hypothesis on who people consider the GOAT by various demographic factors like age, gender, place of residence or favorite sport.

Section 6.) The Exciting Conclusion: 11/13 is less than 6/18.

Using two different approaches, one strictly based on right tail probability and one based on "championship value" we conclude that the Brady-era Patriots did something that is actually slightly more impressive than the Russell-era Celtics. We encourage anyone who pursues this further, especially by exploring some of our questions in Section 5, to support or challenge this result!

Acknowlegements

Our primary sources for data were the excellent websites www.basketball-reference.com, www.baseball-reference.com and www.pro-football-reference.com . In particular, these pages were especially helpful:

https://www.basketball-reference.com/players/r/russebi01.html

https://www.pro-football-reference.com/players/comeback.cgi?player=BradTo00

References

[1] Malik, Blake, "No More Arguments: Tom Brady is the GOAT", downloaded from https://nsucurrent.nova.edu/2021/03/02/no-more-arguments-tom-brady-is-the-goat/ , 2021

[2] Quinn, Justin, "WATCH: Making the case for Bill Russell as the G.O.A.T. on his 87th birthday, downloaded from https://celticswire.usatoday.com/2021/02/12/nba-boston-celtics-bill-russell-greatest-of-all-time/

[3] Author 1, Blinded

[4] Complex Sports Discussion, downloaded from https://www.complex.com/sports/2018/09/the-greatest-of-all-time, 2018.

[5] Bruce, Graeme, "The Greatest Athlete of All-Time According to Americans", downloaded from https://today.yougov.com/topics/sports/articles-reports/2021/04/20/greatest-athlete-all-time-according-americans , 2021.

[6] Deshpande, Sameer K. and Jensen, Shane T. "Estimating an NBA player's impact on his team's chances of winning" Journal of Quantitative Analysis in Sports, vol. 12, no. 2, 2016, pp. 51-72. https://doi.org/10.1515/jqas-2015-0027

[7] Goldsberry, K. and E. Weiss. 2013. "The Dwight Effect: A New Ensemble of Interior Defense Analytics for the NBA." in *Sloan Sports Analytics Conference*, pp. 1–11.Search in Google Scholar

[8] Sabin, R.. (2021). Estimating player value in American football using plus–minus models. Journal of Quantitative Analysis in Sports. 17. 10.1515/jqas-2020-0033.