Classifying GOATs (like Brady, Russell and Ruth) by Measuring Their Tails.

Rick Cleary, Babson College
Steve Miller, Williams College

rcleary@babson.edu, sjm1@williams.edu

MathFest: August 5, 2021
BIG PICTURE SPORTS QUESTIONS...

-Who is the GOAT (Greatest Of All Time) in a particular sport?

-Who is GOAT of GOATs across sports?

We recognize that this is not a well posed question ... but fans and media try to answer it. To do so they make mathematical and statistical arguments that lead them to particular metrics ... sometimes without even realizing it!
Examples abound...

- A student in the school paper at Nova Southeastern University makes the case for Tom Brady in football: TomBradyGOATArgument

- Justin Quinn in USAToday says it’s Bill Russell in basketball: BillRussellGOATArgument

- And at Quora.com Mike Berard makes the case for Babe Ruth: BabeRuthGOATArgument

The Brady and Russell arguments are largely team oriented, while Ruth’s case is more about his individual excellence.
OUR GOAL TODAY...

This is a *preliminary* report on our work to date. We want to:
- Show that metrics matter.
- Give examples of ways that assumptions lead to metrics.
- Choose explicit metrics first and use them to evaluate ‘something like’ a GOAT argument; maybe a ‘best teammate’.

Note: The ideas here can be applied in teaching many applications besides sports. Consistent with ideas in social choice, finance, economics and other fields. Can adjust the technical level to be anything from a first year seminar to a capstone project!
METRICS MATTER - Who’s IN first?

How can we tell who the GOAT if we can’t even decide GORN (Greatest of Right Now)!

Consider June 4, 2018. Boston Red Sox were 41-19, winning percentage .683. New York Yankees were 37-17, winning percentage .685.

ESPN *correctly* reports Red Sox in first (HURRAH!) since they are one “game ahead.”

GOOGLE reports Yankees in first (BOOOO!) since they have higher winning percentage.

<table>
<thead>
<tr>
<th>American League</th>
<th>W</th>
<th>L</th>
<th>PCT</th>
<th>GB</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EAST</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boston Red Sox</td>
<td>41</td>
<td>19</td>
<td>.683</td>
<td></td>
</tr>
<tr>
<td>New York Yankees</td>
<td>37</td>
<td>17</td>
<td>.685</td>
<td>1</td>
</tr>
<tr>
<td>Tampa Bay Rays</td>
<td>28</td>
<td>30</td>
<td>.483</td>
<td>12</td>
</tr>
<tr>
<td>Toronto Blue Jays</td>
<td>26</td>
<td>33</td>
<td>.441</td>
<td>14.5</td>
</tr>
<tr>
<td>Baltimore Orioles</td>
<td>17</td>
<td>41</td>
<td>.293</td>
<td>23</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>American League</th>
<th>W</th>
<th>L</th>
<th>Pct</th>
<th>GB</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AL East</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yankees</td>
<td>37</td>
<td>17</td>
<td>.685</td>
<td>1</td>
</tr>
<tr>
<td>Red Sox</td>
<td>41</td>
<td>19</td>
<td>.683</td>
<td></td>
</tr>
<tr>
<td>Rays</td>
<td>28</td>
<td>30</td>
<td>.483</td>
<td>12</td>
</tr>
<tr>
<td>Blue Jays</td>
<td>26</td>
<td>33</td>
<td>.441</td>
<td>14.5</td>
</tr>
<tr>
<td>Orioles</td>
<td>17</td>
<td>41</td>
<td>.293</td>
<td>23</td>
</tr>
</tbody>
</table>
Or ... Cross country example

Order of finish:

“Invitational” Scoring:  A wins!!
A: 1+4+5+15+17 = 42; C: 6+7+9+10+11 = 43, B: 2+3+8+14+16 = 43.  (C gets second since their sixth runner beat B’s sixth runner.)

“Dual Meet” Scoring: Three different races:
B beats A 27-28; A beats C 29-30; and C beats B 28-31! Everybody goes 1-1!
WHICH COMES FIRST? THE METRIC OR THE GOAT?

You don’t simply choose the GOAT ... you choose a metric!

Scientific method ... Choose the metric and see who’s the GOAT;

OR

As a fan/writer with deadline for a column... Choose the GOAT, then find a metric!

We will pick some reasonable metrics and see what happens!
Let’s have a vote!

In the chat, tell us who you think is the best candidate for GOAT of GOATs:

BRADY ... NFL

RUSSELL ... NBA

RUTH ... MLB

OTHER ... Provide Name/Sport
Let’s follow up with some data

In the chat, we had votes for the best candidate for GOAT of GOATs:

BRADY ... NFL  
RUSSELL ... NBA  
RUTH ... MLB  
OTHER ... Provide Name/Sport

AN ALTERNATE PLACE WE CHOSE TO START: Who is the BOAT (Best Of All Teammates)? Measured by team success relative to league size/quality/playoff format. Preliminary research suggests Brady and Russell is the place to start looking.
The BOAT must ... Get to the playoffs!

Under simplest assumptions, with goal of ‘make the playoffs’. Each team equally likely to qualify each year, years are independent, so a binomial model.

**Brady:** 18 times to playoff in 20 year career in 32 team league with 12 playoff qualifiers. Appearances ~ binomial (20, 12/32).

$$P(18 \text{ or more in 20 years}) = 3.02 \times 10^{-9} = .0000000302.$$
Brady vs. Russell playoff appearances

**Brady:** 18 times to playoff in 20 year career in 32 team league with 12 playoff qualifiers. Appearances ~ binomial (20, 12/32).

\[ P(18 \text{ or more in 20 years}) = 3.02 \times 10^{-9} = .00000000302. \]

**Russell:** Thirteen for thirteen in making playoffs ... but in small league where more than half of teams made playoffs!

\[ P(13 \text{ for 13 in making playoffs}) = 6.29 \times 10^{-3} = .00629. \]

A point in Brady’s favor here!
But if we do titles instead of appearances:

Under simplest assumptions: Each team equally likely to win each year, years are independent, so a binomial model:

**Brady:** Seven titles in 20 years in 32 team league.

\[ X \sim \text{binom}(20, 1/32) \]

\[ P(X \text{ at least 7}) = 1.49 \times 10^{-6} = 0.00000149. \]

**Russell:** Eleven titles in 13 years in (approx.) 10 team league.

\[ X \sim \text{binom}(13, 1/10) \]

\[ P(X \text{ at least 11}) = 6.44 \times 10^{-10} = 0.000000000644. \]

Russell looks better here ... again, metrics matter!!
BUT Brady played GAMES; Russell SERIES

Horizontal axis: \( P(\text{stronger team wins any one game}) \).

Vertical axis: \( P(\text{stronger team wins series of 1,3,5,7 games}) \) under independence.
BUT Brady played **GAMES**; Russell **SERIES**

Interpretation: Series help better teams avoid upsets. Also Brady had to win 3-4 games while Russell usually played two series. Does that make up the difference? IT MIGHT!

<table>
<thead>
<tr>
<th></th>
<th>Years</th>
<th>Titles</th>
<th>Championship Conferences</th>
<th>Playoffs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brady/#2*</td>
<td>83</td>
<td>175</td>
<td>143</td>
<td>156</td>
</tr>
<tr>
<td>Brady/next five</td>
<td>119</td>
<td>184</td>
<td>185</td>
<td>167</td>
</tr>
<tr>
<td>Russell/#2*</td>
<td>65</td>
<td>110</td>
<td>109</td>
<td>93</td>
</tr>
<tr>
<td>Russell/next five</td>
<td>83</td>
<td>157</td>
<td>122</td>
<td>116</td>
</tr>
</tbody>
</table>

**Ratios: Brady and Russell to Top Competitors**
The log 5 rule (due to Bill James)

Suppose team A wins games with probability $p$, and team B wins with probability $q$.

A good estimate of $P(A$ wins a game vs. $B)$ is given by

$$\frac{p(1-q)}{p(1-q)+(1-p)q} = \frac{p-pq}{p+q-2pq}$$

Example: In a playoff round we might have $p = .8$ and $q = .6$

So $P(A$ wins a game vs. $B) = (.8 - .48) / (1.4 - .96) \approx .727$.

(See SJM’s paper at https://web.williams.edu/Mathematics/sjmiller/public_html/399/handouts/Log5WonLoss_Paper.pdf)
A Log-5 Adjustment for series vs. games

Team A, P(win)=p, plays B with q=0.6. Maximum ratio is about 1.6. Two series.

James Log-5 Adjustment: Playing a team that wins 60%
A Log-5 Adjustment for series vs. games

Team A, P(win) = p, plays B with q = 0.6. Maximum ratio is about 1.9. Three series.

James Log-5 Adjustment: Playing a team that wins 60%

Prob win three series
So here’s an argument for Brady...

RUSSELL ... 11 titles.

BUT NFL titles might be about 1.8 times more difficult because of the games/series issue!
So here’s an argument for Brady...

RUSSELL ... 11 titles.

BUT NFL titles might be about 1.8 times more difficult because of the games/series issue!

So BRADY’s seven titles might be worth about:

\[7 \times 1.8 = 12.6 \text{ RUSSELL titles!}\]
-Some next steps ...

We have begun considering (and will recruit students to help us consider):

- More sophisticated metrics with playoff wins models (Poisson vs. binomial);

- Championship round success (proponents of Michael “six for six” Jordan and Joe “four for four” Montana

- Similar argument for longevity vs. high peak. Is four title in eight year career more or less impressive than four in 20?

- Influence by sport (Russell one of five, on floor 80% of time; Brady one of eleven, of field 40% of time ... but Brady key in all of those plays while Russell might go some time without a touch.)

- We’ll think of more!
One more vote ...

Based on our discussion, please use the chat to vote again for BOAT:

BRADY    RUSSELL    RUTH    OTHER

Did we change any minds?
(We did change the question!)
THANKS for your attention ...

-Any questions please contact us at rcleary@babson.edu or sjm1@williams.edu
Slides below here are for another day...
## Players to consider... NFL

<table>
<thead>
<tr>
<th>Player</th>
<th>Years</th>
<th>Titles</th>
<th>Finals</th>
<th>Semis</th>
<th>Playoffs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tom Brady</td>
<td>2002-21</td>
<td>20</td>
<td>7</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>Joe Montana</td>
<td>1980-94</td>
<td>15</td>
<td>4</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Terry Bradshaw</td>
<td>1971-84</td>
<td>14</td>
<td>4</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Jerry Rice</td>
<td>1986-05</td>
<td>20</td>
<td>4</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Steve Young**</td>
<td>1986-00</td>
<td>15</td>
<td>3</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Roger Staubach</td>
<td>1970-80</td>
<td>11</td>
<td>2</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>John Elway</td>
<td>1984-99</td>
<td>16</td>
<td>2</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Adam Vinatieri</td>
<td>1997-20</td>
<td>24</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>
Players to consider... NBA

<table>
<thead>
<tr>
<th>Player</th>
<th>Seasons</th>
<th>Yrs</th>
<th>Titles</th>
<th>Final</th>
<th>Semis</th>
<th>Playoffs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bill Russell</td>
<td>1958-70</td>
<td>13</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>Michael Jordan</td>
<td>Various</td>
<td>15</td>
<td>6</td>
<td>6</td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td>Sam Jones</td>
<td>1958-69</td>
<td>12</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>K Abdul-Jabbar</td>
<td>1971-90</td>
<td>20</td>
<td>6</td>
<td>10</td>
<td>14</td>
<td>18</td>
</tr>
<tr>
<td>LeBron James</td>
<td>2004-21</td>
<td>18</td>
<td>4</td>
<td>10</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>Tom Heinsohn</td>
<td>1957-65</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Magic Johnson</td>
<td>Various</td>
<td>13</td>
<td>5</td>
<td>9</td>
<td>10</td>
<td>13</td>
</tr>
</tbody>
</table>
GOAT METRICS IN TEAM SPORTS ... 

- Individual statistics?
  - Counting stats like Points, Touchdown Passes, Home Runs, Goals, Assists
  - Derived stats like Wins Above Replacement, Adjusted Plus/Minus

- Team statistics?
  - Titles won (consider time periods, league size)
  - Consistent excellence (high winning percentage over a long time.)
Measuring GOATS by their tails...

Principle: What a GOAT does is accomplish something that is *least likely* among a set of candidates using a particular metric. There are many variables to consider when computing these probabilities, but as an exercise...

Positives:
- Great modeling project that requires careful consideration of inputs and assumptions.
- Adaptable; can be used for both team and individual performance.
Measuring GOATS by their tails...

Principle: What a GOAT does is accomplish something that is least likely among a set of candidates. There are many variables to consider when computing these probabilities, but as an exercise...

Positives:

- Great modeling project that requires careful consideration of inputs and assumptions.
- Adaptable; can be used for both team and individual performance.

Negatives:

- Sort of like a p-value. Why right tail only?
- Almost certainly not definitive.
The “What Abouts...”

- What about teammates? Russell played 80% of the time and was one fifth of the team at any time. Brady played 40% of all plays and was 1/11 of his team at any time.

Another one we might not have time to explore in talk
The “What Abouts...”

- What about influence? Russell played 80% of the time and was one fifth of the team at any time. Brady played 40% of all plays and was 1/11 of his team at any time.
- What about league parity? The NBA in Russell’s era had teams that were never truly competitive.
The “What Abouts...”

- What about influence? Russell played 80% of the time and was one fifth of the team at any time. Brady played 40% of all plays and was 1/11 of his team at any time.
- What about league parity? The NBA in Russell’s era had teams that were never truly competitive.
- What about quality of teammates? (And many more...)

So stay tuned: We hope to attract students to build these more sophisticated interpretations!