**Evaluating RPI: Ranking College Basketball Teams**

Saj Sri-Kumar and Sarah Rosston

The Ratings Percentage Index, or RPI, is one of the most important tools that the NCAA Division I Men's Basketball Tournament Selection Committee uses to determine which teams should be selected for the at-large bids into the postseason tournament, popularly known as March Madness. The ranking system weights a team's winning percentage (25%), opponents' winning percentage (50%), and opponents' opponents' winning percentage (25%). The system has been criticized on many fronts, most notably for the low value it puts on a team's own winning percentage--75% of a team's RPI is completely out of its control and only is dependent on the performance of their opponents. Furthermore, the weighting was created with no attention paid to the measure's ability to serve as an effective measure of performance or its statistical meaning. Despite the criticism, it remains used by the selection committee because it, unlike many other ranking systems, does not include margin of victory in the calculation. Although margin of victory has been shown to be one of the best indicators of future success (especially postseason success), the NCAA does not condone its usage to rank teams as it is afraid that allowing its use would encourage teams to "run up the score" by scoring high numbers of points at the end of the game even if the existing margin is sufficient to assure victory.

In this project, we considered many alterations in an attempt to improve the metric. We examined ESPN's newly created "Basketball Power Index", in an attempt to find if it was better at predicting success in the tournament. We considered the same inputs that are present in the RPI, but with different weightings, then created additional formulae that

included margin of victory as well as incorporating our own ranking system of conferences in an attempt to find an alternative measure of schedule strength.

**Basketball Power Index**

In 2012, ESPN created a new metric to compete with the RPI and other ranking systems, called the "Basketball Power Index", or BPI. In addition to the winning percentage and strength of schedule measures, the BPI includes numerous other factors. First, it includes margin of victory, but also corrects to make sure that there are diminishing returns to blowouts. In other words, a win by 25 points is much better than a win by 2 points, but not much worse than a 40-point win. Furthermore, it also accounts for games in which a team's key players were missing, as those games were deemed not to be representative of a team's potential going forward (assuming that those players would be healthy enough to return).

To evaluate the metric's success, it is useful to examine how accurate the model would have been in previous years. Although the exact formula is not public--and is likely too complicated to be able to used with publicly available data--ESPN ran its own calculations on tournaments between 2007 and 2012 and found that BPI would have correctly predicted 66% of matchups correctly (compared to 61% for RPI). While that is only a modest improvement, it was much more successful at predicting how far into the tournament a team would go (i.e. which round they would be eliminated in), correctly predicting 3 of 7 national champions (compared to zero for RPI), and 12 of 28 final four teams (compared to only 6 for RPI).

**Our Models**

While we were not able to use some of the advanced data techniques that ESPN was able to use, due to lack of data as well as unfamiliarity with some of the statistical techniques used, we were able to create models using basic data such as win percentage strength of schedule, and our conference rankings (detailed below). Similarly to ESPN, we compared the models to tournament data, attempting to improve on RPI as a way to predict how far into the tournament a team would travel. To account for the fact that difficulty advancing in the tournament does not increase linearly (i.e. it is more than 2 times harder to get out of the second round of the tournament than it is to make it out of the first round), we used an exponential function for rounds, and used regression to compare that to our models.

**Conference Rankings**

Before raking teams individually, we attempted to rank conferences as a whole. In college basketball, most teams play teams outside of their conference for the first few months of the season, and then starting in January, play teams mainly in their conference. The nonconference portion of the season provides a critical window with which we can compare the conferences themselves--as teams spend the second half of their season playing teams within their conference, it is critical to have a measure of how strong that competition is. In other words, while a team may be able to beat other teams in its conference, that information is hard to gauge in the wider context of all of the different conferences. By ranking the conferences themselves, we can, in effect, judge the "quality" of those wins rather than just marking them as wins in an abstract sense. Ideally, we would have records in head-to-head matchups between conferences, and use something like

PageRank to distribute weights, but we only found non-conference win percentage, so we ranked conferences on their average non-conference win percentage.

**Results**

Using linear regression, all of our models had significance in predicting tournament success. We used RPI as a baseline metric, and several of our models outperformed RPI

```
Call:
lm(formula = 2^Round.lost ~ RPI.y - 1, data = merged)

Residuals:
    Min      1Q  Median      3Q     Max
 -7.072  -6.028  -4.978  -2.469 119.298

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
RPI.y   14.296      3.846   3.717 0.000435 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 17.96 on 62 degrees of freedom
  (176 observations deleted due to missingness)
Multiple R-squared:  0.1823,    Adjusted R-squared:  0.1691
F-statistic: 13.82 on 1 and 62 DF,  p-value: 0.0004346
```

Fig. 1. : Modeling tournament progression with RPI

The next model we used assumes that the median non-conference win percentage is a good indicator of the strength of a conference, which we used to weight the conference win percentage. We believe this works better than balancing an individual team's conference and non-conference win percentage because of the varied strength of non-conference schedules.

```
Call:
lm(formula = 2^Round.lost ~ Conf.Win.pct:conf.median - 1, data = merged)

Residuals:
    Min      1Q  Median      3Q     Max
 -9.221  -6.304  -4.478  -1.569 118.648

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
Conf.Win.pct:conf.median   18.389      4.722   3.895 0.000243 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 17.8 on 62 degrees of freedom
  (176 observations deleted due to missingness)
Multiple R-squared:  0.1966,    Adjusted R-squared:  0.1836
F-statistic: 15.17 on 1 and 62 DF,  p-value: 0.0002435
```

Fig. 2: Using conference performance against other conferences to weight conference win percentage. P-values and R^2 were better than RPI

Next, we use the pythagorean theorem of sports, which gave us better results than RPI, but not as good as the conference-ranking model.

```
Call:
lm(formula = 2^Round.lost ~ I((Team.points^2 + Opponent.Points^2)/(Team.points^2)) -
    1, data = merged)

Residuals:
    Min      1Q  Median      3Q     Max
 -6.854  -6.179  -4.358  -1.686 120.223

Coefficients:
                                                    Estimate Std. Error t value
I((Team.points^2 + Opponent.Points^2)/(Team.points^2))    4.382      1.265   3.463
                                                    Pr(>|t|)
I((Team.points^2 + Opponent.Points^2)/(Team.points^2)) 0.000974 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 18.18 on 62 degrees of freedom
  (176 observations deleted due to missingness)
Multiple R-squared:  0.1621,    Adjusted R-squared:  0.1486
F-statistic: 11.99 on 1 and 62 DF,  p-value: 0.0009742
```

Fig. 3. Using the Pythagorean Theorem of Sports

Unsurprisingly given ESPN's data and models, BPI outperformed RPI and all of our models.

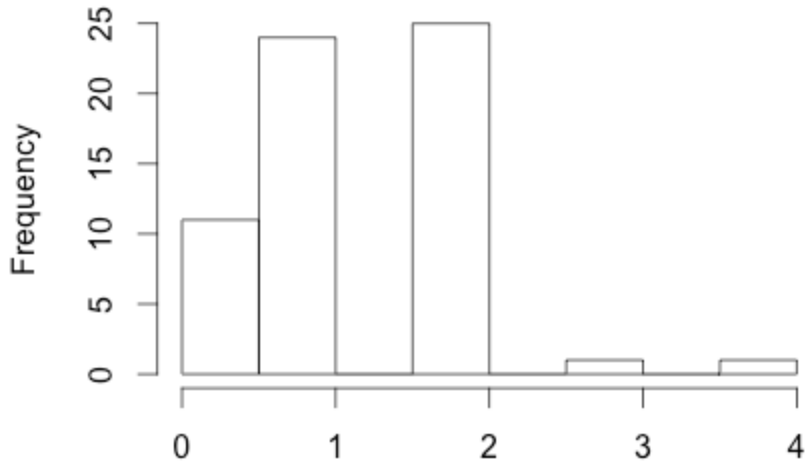
```
Call:
lm(formula = 2^Round.lost ~ BPI - 1, data = tournmanentNew)

Residuals:
    Min      1Q  Median      3Q     Max
 -8.037  -6.396  -4.932  -2.462 118.131

Coefficients:
    Estimate Std. Error t value Pr(>|t|)
BPI  0.11991    0.03054   3.927 0.000222 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17.88 on 61 degrees of freedom
  (176 observations deleted due to missingness)
Multiple R-squared:  0.2018,    Adjusted R-squared:  0.1887
F-statistic: 15.42 on 1 and 61 DF,  p-value: 0.0002216
```

Fig. 4. BPI produced the best results of all of our models

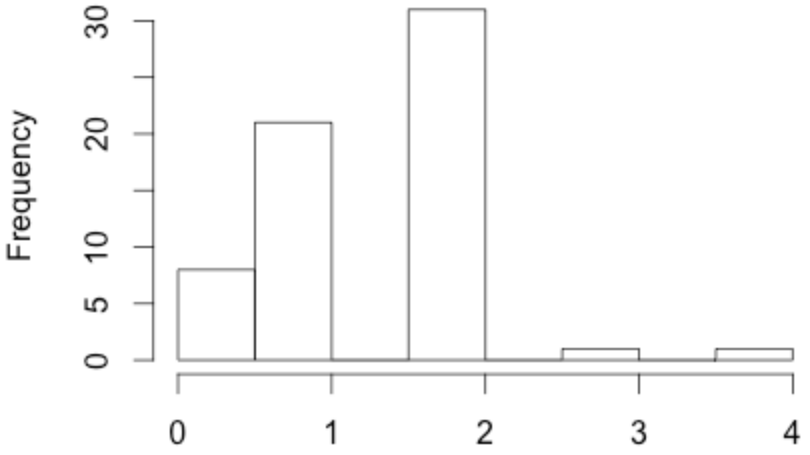
When we tried reweighting the components of RPI, we found that the highest R^2 value we could get was 0.1881278, with weighting only a team's win rate.

None of the models were great at predicting when teams would exit the tournament, and all skewed to high, which a model that imposed better limits on how many teams could reach each round might have fixed. The distribution of misses for each model are shown below:

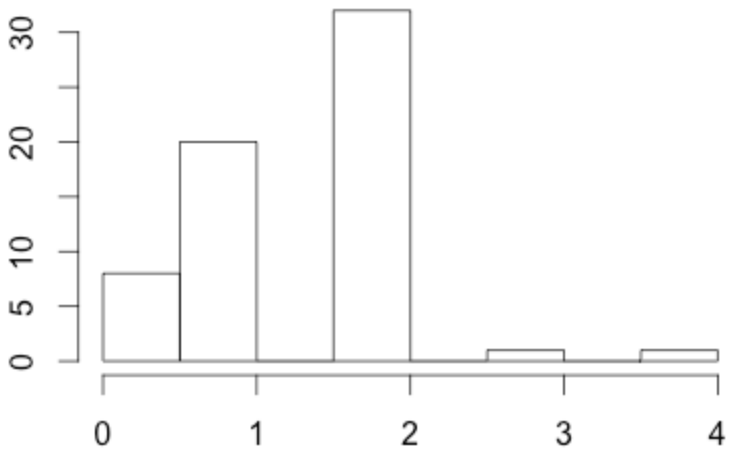
**Distribution of Difference between Conference Model and Actual Results**



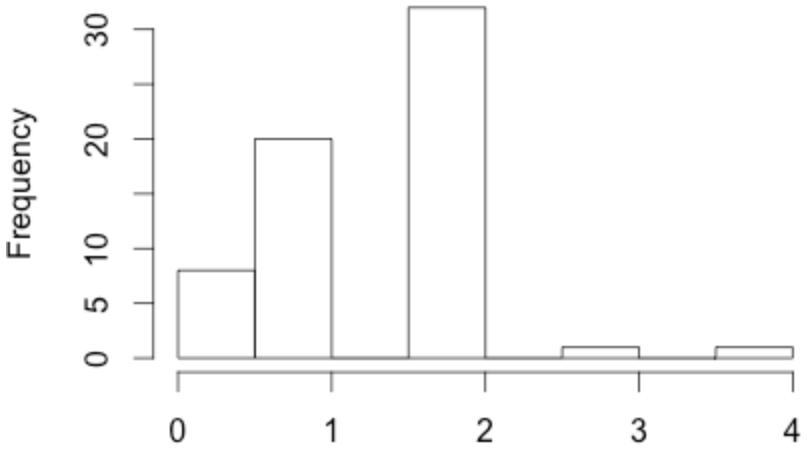

**Distribution of Difference between BPI Model and Actual Results**



**Distribution of Difference between RPI Model and Actual Results**




**Distribution of Difference between Pythagorean Model and Actual Results**




The graphs are all relatively similar, which probably comes from both upsets and more qualitative data that was not included in the models.

**Error Analysis**

While our formulas were better at predicting last year's tournament than RPI, a lack of data hindered our efforts. Summary statistics for teams were easy to find,but a list of

teams, opponents, and scores for the season was not. We found information about win-loss records in and out of conference play, and points score but we could not find data about head-to-head matchups. This lack of data prevent us from using models like Bradley-Terry, or creating a better ranking of conferences. With this added data, we think we could have created a model that would have outperformed RPI more significantly.

**Conclusion**

Even with significantly less data than the March Madness Selection Committee, we were able to come up several models that significantly outperform RPI. All of our models predict tournament success more accurately than RPI, but less accurately than ESPN's BPI rating. Although we probably could have drawn better conclusions with more head-to-head matchup data, we could easily find better summary statistics than the ones used in RPI. Our performance indicates that RPI could benefit significantly from using more data like conference power rankings and score differential.