Ratings methodology update

[Ken Pomeroy](#) | 10.04.16

(For some words on the previous iteration of the ratings and what the columns on the ratings page mean, go [here](#).)

You may have noticed that the ratings pages look a little bit different than they used to. The Pyth column has been changed to AdjEM (adjusted efficiency margin) and likewise, strength of schedule and conference strength measures have been converted to the new scale.

The main benefit of the change is that the team rating means something that is more easily understood by humans. The problem with the "Pythagorean winning percentage" was (a) it was a mouthful, and (b) you can't easily compare the relative strengths of teams. For one thing, the scale isn't linear. The difference between .98 and .97 is not the same as the difference between .52 and .51 in terms of team strength. Furthermore, what do those numbers mean anyway? They do have meaning – expected winning percentage against an average D-I team – but when comparing two teams it's not very clear what the difference means.

AdjEM is the difference between a team's offensive and defensive efficiency. It's simple subtraction. Even your dog can do it. It represents the number of points the team would be expected to outscore the average D-I team over 100 possessions and it has the advantage of being a linear measure. The difference between +31 and +28 is the same as the difference between +4 and +1. It's three points per 100 possessions which is much easier to interpret. This measure also makes the SOS and average conference strength numbers less mysterious.

In order to make this work properly the method to produce the ratings needed to be tweaked as well. All past ratings reflect these changes.[1] The rankings evolution for each team have not been changed either, but I expect to eventually update that at some point.

(The remaining words are devoted to the tedium of ratings architecture. Reader discretion is advised.)

Previously, the adjusted efficiencies on offense and defense were computed using principles borrowed from Dean Oliver. Essentially, the expected offensive efficiency for a team was the product of its season-long offensive efficiency and its opponent's season-long defensive efficiency. If Team A's offensive rating is 110% of the national average and Team B's defensive rating is 110% of the national average, then Team A's offense when playing Team B would be expected to be 110 x 110 or 121% of the national average.

In the new system, the effects of the two competing teams are considered to be additive rather than multiplicative. If Team A's offensive efficiency is 10% higher than the national average and Team B's defensive efficiency is 10% higher than the national average, then Team A's offense when playing Team B would be expected to be 20% higher than the national average.

I'm not sure how basketball really works, but my hunch is that it's probably closer to additive than multiplicative. For fairly normal teams, the distinction is nearly irrelevant. But at the extremes, it can matter. If Team A has an offense than is 120% of the national average and Team B has a defense 80% of the national average, it makes intuitive sense that Team A's offense should be exactly average when it plays Team B. In the multiplicative framework Team A's offense would be expected to be 96% of the national average, implying that a great defense is better than a great offense, which seems to conflict with reality.

Another benefit of the additive model is that it allows us to more easily separate the influence of offense and defense in making predictions so one could weight offensive rating more heavily. Preliminary research into this is promising, but for this season, offense and defense will be weighted equally until a more extensive investigation can be conducted.

While offense and defense are still rated independently, the performance of each is now evaluated against the national average in efficiency on the date the game was played.[2] Opponent quality being equal, it's more impressive to post 1.4 points per possession in a November game than a March game since average efficiency tends to rise during the season, and the new system accounts for that. Again, this isn't a huge thing, but while I was tinkering with code, it felt like something worth adding.

There have been some other changes to the system as well. The weighting coefficients to handle recency and game importance have been changed. Essentially, recency is less important than it used to be and game importance is less sensitive to margin and opponent than it used to be.

This serves to provide a bit more spread in the ratings. Previously, predictions involving teams of significantly different ability would give the underdog too much credit. And the system tended to give teams that dominated weak conferences a bit too much credit. Those issues should be improved now.

In addition, dropping the pythagorean winning percentage as the team measure means that the days of estimating win probabilities using log5 are over. That method also [has some problems in mismatches](), even in a well-calibrated system. Now win probabilities will be derived directly from the predicted margin of victory.

There are some other useful changes as well. The strength of schedule

rating is now more fair. After a through search, I'm now using an implementation of [Jeff Sagarin's WIN50 method](#).[3] The SOS figure represents the strength of team that would be expected to win half its games against the team's schedule. It is handy because it minimizes the effect of outliers on the SOS calculation while allowing everyone's SOS value to be compared on the same scale.

If a team plays mostly tough opponents, then the SOS rating isn't as sensitive to the quality of the bad teams it plays. Whether Texas played Central Connecticut instead of UTSA last season wouldn't have changed its SOS much. Flip this principle for a team that has mostly bad opponents on its schedule. Mostly this won't have a big impact, but under the previous method, which used a simple average of its opponents' ratings, whether you played the 350th- or 351st-best team in 2013 would have had far-too-serious implications on one's SOS.

The same method is also used to rate conferences. A conference's rating is the strength of team that would be expected to go .500 against a round robin schedule. This, too, reduces the effect of outliers. For instance, the Mountain West jumps from 12th to 10th in 2015 as the effect of #349 San Jose State is reduced.

One could argue that the baseline team should be .750 or .900 instead of .500 since maybe we mainly care about the quality of the best teams in a conference. I would not like to spend much energy rebutting that philosophy but I do like that the conference ratings and SOS ratings are on roughly the same scale. Only in cases where conferences are very closely rated will the choice of baseline make a difference.

Finally, there is the issue of home court advantage. For this iteration of the ratings I am using a flat 3.75 points for every game. It's probably a

little higher than reality since that value is what best calibrates predictions of the past 15 seasons, and we know that home court advantage has been on a subtle decline in recent years. It's a high priority of mine to implement site-specific home-court advantage values by next season. It isn't something that matters much in the long run, but it's kind of a neat thing to have and hopefully I can discuss this more during the season.

References

^1	Past predictions do not reflect these changes.
^2	Previously, the season-long national average of efficiency was used for all games played.
^3	Jeff probably didn't invent this, so I apologize to whoever did, but he is the most famous person using it.