Is it Possible to Objectively Generate the Rankings produced by the College Football Playoff Committee?

John A. Trono
Saint Michael's College
Colchester, VT
Tech report: SMC-2016-CS-001

Abstract

Several methodologies are described that attempt to objectively generate the same ranking that is produced by the committee who determines which teams will be invited to the NCAA college football playoff (CFP). Two models have been discovered that have correctly predicted the opponents in both first round contests in the first two years of the CFP. These models' rankings were also reasonably close throughout the entire top 25 list as announced by said committee.

Introduction

Many collegiate sports conclude their seasons with a tournament to decide which team will be recognized as the national champion (NC) that year. Sports like baseball, basketball, hockey and soccer (to name a few) will hold conference playoffs, at the end of the regular season, to determine which teams will compete in such tournaments, though many sports allow for additional teams to be invited, to compete as 'at large' selections, as well. However, the National Collegiate Athletic Association (NCAA) division one football season had historically concluded with only its own bowl game pageantry, generating controversy concerning which team should be recognized as the best that year, though lower divisions had instituted single elimination tournaments to crown their champions. For example, starting in 1973, the division two NC (in football) was determined by a single elimination tournament of 8 teams, which increased to 16 in 1988, and then to 24 – with 8 teams receiving a first round bye – in 2015. The division three NC was also crowned after 4 teams were invited in 1973 and 1974; 8 teams were invited for the next 10 years, which was increased to 16 teams in 1988, 24 in 1999 and 32 teams in 2005 – requiring a total of five weeks to determine the NC.

Two major groups of people (selected sportswriters, and chosen coaches) have been casting votes as to who they think are the top 25 division one football teams at year's end as well as after each week's games. Given the subjective nature of such polls, the sentiment regarding which team was the 'consensus NC' has been different in those two, post bowl game polls at the end of many seasons. There will always be disagreements as to which team **was** the best that year, or which one had the 'best' season, etc. However, because certain bowl games maintained invitation obligations with certain conferences, there were too many times when the perceived top two teams would be 'required' to participate in different bowl games. 1990, '91 and '94 are just a few of the years which helped fuel the creation of an improved resolution to this problem because the question regarding which of those top two teams was better was not previously decided as it should be: on the field.

To rectify this situation, the Bowl Championship Series (BCS) began in 1998. Its purpose was to create a system where the two best teams would meet, with the winner being crowned as that year's NC. During the 16 years when the BCS methodology selected the championship game's competitors, there were several occasions where it was hard to narrow the field down to just two teams when it appeared that there were three (or four, or more) teams whose resume that year was worthy of being selected to compete for the national championship.

In 2014, this small field was expanded to include the top four teams, as chosen by the college football playoff (CFP) committee, and this seems to be an improvement over the BCS approach, though some will argue that the top eight teams might have earned the right to compete for the NC. Given the extreme physical nature of football, leading to the lengthy recovery time between games (i.e. a week) – for physical safety reasons perhaps – and the time of the year when the division one season concludes, it is unlikely that a tournament similar to those for division two and three will ever be realized for division one football.

Even given the limitations inherent in the current four team playoff, not many would argue against this as an improvement over its predecessors. With the current format, any NC must have defeated two very outstanding teams to have earned the accolades for becoming the recognized best team that season. However, if there is no objective methodology to select the four competitors, it is quite likely that any decision will result in some controversy, with regards to those worthy teams who were not chosen for this playoff.

The Final Two Weeks in the 2014 Regular Season

The CFP committee members were selected prior to the 2014 season. This committee met each week, starting after week #9 (of the 15 week season), to produce their top 25 team ranking, given the evidence – as displayed on the field – up to that point in time. After week #14, the top six teams were: 1) Alabama (11-1), 2) Oregon (11-1), 3) TCU (10-1), 4) Florida State (12-0), 5)

Ohio State (11-1), and 6) Baylor (10-1). Many football fans thought that Baylor should've been ranked above TCU, given they had already beaten them. However, TCU had defeated a strong Kansas State team (9-2), along with West Virginia (6-5) – the only team that Baylor had lost to. Given that Baylor was playing Kansas State in week #15, knowledgeable fans understood that a Baylor win in the final week would probably propel them above TCU, regardless of how many points TCU won by in their final game against Iowa State (2-9).

Since all the major conference championship games were also played in that final week, the top four spots were certainly 'up for grabs'. Three of the top four teams did win their conference's championship game. However, #5 Ohio State's victory in the Big-10 championship game (led by their third string quarterback), over Wisconsin (10-2) by a score of 59-0, helped to convince the CFP committee to rank the Buckeyes as the #4 team. This produced many harsh criticisms of the CFP committee, especially from those two co-champions (Baylor and TCU) of the Big 12 conference – which did not schedule a championship game in 2014.

To determine if Ohio State (along with the three teams ranked above them) had actually performed to the level to be included in the CFP's three game tournament, it seemed worthwhile to attempt to quantify how successful each team's season was in 2014. The CFP committee examined each team's entire 'body of work' that year, and it probably gave strong consideration to 'quality wins', i.e. wins over other (perceived) strong teams, and the committee would probably look poorly on losses – especially to teams with less than exemplary records. In an effort to mimic the result generated from the CFP committee deliberations, it seemed plausible that teams could be grouped (or clustered) with other 'like teams', and then a win over any team in a group could be assigned a certain point value, with the four teams accumulating the most points being recognized as having earned their invitation to the CFP.

One guideline for determining the size of these aforementioned groups was to make them smaller in size as the teams in that group had 'performed better' than the preceding group; likewise, the points earned would increase for wins over teams in these smaller, more selective/impressive groups. Since there were 128 division one teams in 2014, an exponential decrease in group size could easily produce a final group of size four. The original idea was that wins against the bottom half of the 128 teams (64 to be exact) would only earn 1 point, but 2 points would be garnered for defeating a team in the next group (of 32 teams), 4 points for a victory over one of the next 16 teams, 8 points for the next group of 8 teams, 16 points for a win over the next group of 4, and 32 points for a win over a top four team.

Unfortunately, there were two obvious shortcomings with these initial details. The first was that teams might be rewarded too highly for wins against top teams, when applying the aforementioned group sizes (and point values). This could lead to some perpetual back and forth movement between the top teams in the last two groups (of four teams) if the determination of which teams belonged in which groups would be repeated until convergence had occurred. For example, given some initially assigned grouping, let's say that team A had defeated team B, who currently resided in the top group of 4; that win just might move team A into the top four group. And, after further review, team B might also have dropped into the next lower group of four. Meanwhile, team C could've defeated team A, and could now appear to 'belong' in the top four group as well. This circular reassignment pattern could continue indefinitely, i.e. team A is promoted to the top four, but is then replaced by team C, followed by team B (who won its game against Team C) being promoted back into the top group of four teams *ad infinitum* (with the 'defeated teams' also moving downward in said groupings during each reevaluation).

The other concern was that all teams do not play the same number of games in a season. One way to rectify that, and to not reward teams for padding their schedules with 'inferior' competition, was to allocate zero points for a win over a team in the lowest (and largest) group.

Experiencing a loss will also cause a number of points to be subtracted from a team's current point total, and that number would be the 'inverse' of the points associated with wins over teams in that group, with a +1 added to the number to be subtracted so that even a loss to a top team still earns a small penalty. For example, a win over a top four team would now earn 16 points, when starting at zero (for the largest group); therefore, a loss to a team in the bottom half incurs a 17 point penalty.

Other Grouping Strategy Alternatives

Before testing the exponential decrease strategy (E), seven other reasonable grouping rules were devised, and all of these are also designated by one letter names: E, W, N, O, F, V, H, and G. The E strategy, as previously presented, was modified slightly – stopping with the smallest (and most selective) group size at eight; this adjustment could help to alleviate the scenario where teams might repeatedly move back and forth between the two groups of size four, as previously described. The main guideline when creating such grouping strategies was that each subsequent group size should be smaller than its predecessor, thereby essentially increasing the level of selectivity as each prospective group shrinks in size.

Several of these grouping strategies are similar to the fraction-based approach utilized in E that is made up of decreasing group sizes of 1/2, 1/4, 1/8, 1/16, and (the final) 1/16 of all teams, respectively. The next, most obvious grouping strategy had to be dropped since 1/2 + 1/3 + 1/4 > 1. However, 1/3 + 1/4 + 1/5 + 1/6 equals 0.95, so in this case (W), the last group size is very selective, and is comprised of only the top 1/20th of all the teams.

Given the extreme size discrepancy between the last two groupings in W, the next grouping (N) also patterned itself as an abbreviated application of Zipf's law, using the decreasing group sizes of 1/4, 1/5, 1/6, 1/7 and 1/8, which would leave almost exactly 1/9 in the smallest group. Strategy O simply uses only odd denominators (from 3 down to 13) for the fractional group sizes, with the final group being comprised of roughly the remaining 1/23 of the teams not already placed into a prior group.

The other four strategies are percentage based, with F being the simplest; it starts at forty percent, and each subsequent group is ten per cent smaller than the preceding group, i.e. 40, 30, 20 and finally 10 percent. Strategy H puts half of the teams in the first two, larger groups – each holding 25 percent of all teams – and then each of the next groups are five percent smaller than the previous group (25, 25, 20, 15, 10, and 5%), whereas strategy V, which also concludes with the smallest group at 5%, starts with 35%, then 30%, then 20%, and the penultimate group of 10%. Finally, strategy G, which stands for the 'Golden Ratio', i.e. $(1 + \sqrt{5}) / 2$, was an attempt to have the relative sizes of adjacent groups to be roughly 0.618 (which is the reciprocal of the Golden Ratio). To have approximately 4% remaining in the smallest group, the first group begins with 37.5% of the teams, and this group size is multiplied by 70% to produce the second group's size; all the remaining group sizes will be 61.8% of its predecessor (rounding when necessary – with the smaller group sizes matching the Fibonacci sequence, in reverse order, i.e. 34, 21, 13, 8 and 5, respectively, when 128 teams are present).

The number of groups for this set of eight strategies ranges from four to seven, with all but two strategies generating five or six groups in total. While it may appear to be the case that there are plenty of similarities between these strategies, only empirically evaluating them will determine how closely they might mimic the results produced by the CFP committee (as some subtle differences in grouping may match the committee's decisions more closely).

Other Point Values Options

Besides the prior, exponentially increasing example (0, 1, 2, 4, 8, 16, 32), there are three other schemes that seemed worth exploring. Q implements quadratic growth, via linearly incremented increases (0, 1, 3, 6, 10, 15, 21), while L is simply a linear function (0, 1, 2, 3, 4, 5, 6). Finally, an increase from group to group that is similar to the Fibonacci sequence was included as well: (0, 2, 3, 5, 8, 13, 21).

Of the 32 possible combinations (for the eight grouping strategies, and, the four point value choices), 24 chose Alabama, Oregon, Florida State and Ohio State as the top four teams: four combinations had them in the same order as the CFP committee, six had Oregon and Florida State in reversed order (as #3 and #2 – but this wouldn't have changed which teams played each other in the 2014 playoff), and another 14 produced various ranking permutations of these four teams. (These results were observed after incorporating a 'reasonable' strategy to order the teams initially, and these point totals were accumulated after only making **one pass** through the scores for that season.) Table 1 illustrates all 32 combinations with: '*' indicating a perfect match; '#' representing the reversed order of Oregon and Florida State; '+' signifying this combination correctly matched who the CFP thought were the top four teams; and '--' meaning that at least one of the CFP committee's top four teams were not present in the list of the top four point values that were generated.

	E	W	N	O	F	V	H	G	Sums
E	--, 7	--, 4	*, 50	--, 3	+, 42	--, 42	--, 3	--, 5	156
Q	--, 9	--, 3	#, 42	*, 50	+, 42	*, 50	--, 5	+, 35	236
L	#, 42	+, 35	+, 35	#, 42	*, 50	#, 42	+, 38.5	+, 35.5	320
F	#, 42	+, 35	#, 42	+, 42	+, 45	+, 38.5	+, 35.5	+, 36	316
Sums	100	77	169	137	179	172.5	82	111.5	

Table 1 – Quantitative model comparison for all grouping/point value combinations (in 2014)

As it seemed desirable to further uniquely quantify each combination's performance, the following rules were employed to produce a number representing the level of matching present, with respect to the CFP committee's ranking. First off, 10 points are awarded for correctly matching the committee's top four teams, plus another 10 points per team for matching exactly the same rank for that team as determined by the CFP. (A perfect match would then be worth 50 points.) If all of the top four teams were ranked correctly, then 6 points are awarded for each team's predicted rank that is off by one position (from where the committee ranked them), 3 points when off by two, and 1 point if off by three (from the CFP ranking). To highlight which combinations did not match all four, a correctly ranked team is only worth 3 points and 1 point otherwise. A tie occasionally is produced as the final result of evaluating some of the aforementioned 32 combinations, and in that case, the point values were averaged. For instance, with HL above, Alabama and Florida State were tied for first, and so the point value calculated with them ranked #1 and #2 was averaged with the point value determined if they were ranked #2 and #1.

In every one of the eight combinations that did not match the CFP committee, all of them did not have Ohio State in the top four; the Buckeye's rank ranged from 5th to tying for 7th in these cases. Three times TCU was #4, and likewise for Arizona; Baylor ended up 4th once, and TCU was tied with Mississippi State for that position in one other combination.

In some sense, it was somewhat surprising that 75% of these 'arbitrarily devised' combinations agreed with the committee's choices for the top four teams, but it was also somewhat reassuring that the CFP committee produced a ranking that was not unreasonable (as many argued). Given the high level of agreement between this simple type of model, and the CFP final ranking, it appeared that the committee had reasonable justification for their final ranking in 2014. (Again, this is not to say that the committee actually, correctly, selected the 'four best teams', just that this **evidence** is now available to support their decision.)

Grouping strategy F had the highest column sum (just barely above V), and point value strategy L was the highest row sum (again, just barely above F), so if one were forced to choose a combination, FL would seem like the 'best' candidate at this time. (Even if the eight combinations, which did not match the top four teams, were reevaluated in relatively the same manner as the other 24 combinations, excluding the ten point bonus, the four affected columns, and two rows, would still have sums below the pair chosen above. For instance, the EQ value of 9 would improve to 30 since it did match three of the top four exactly, etc.) And, it is interesting to note that this combination (FL) did match the committee's top four exactly (as indicated by the 50 in the corresponding cell, in Table 1).

The 2015 regular season

There seemed to be less controversy at the end of the next season (2015) since all four champions, of the five major conferences, with one or fewer losses, appeared in the committee's top four. However, reevaluating the same 32 combinations (of a grouping and point value strategy) produced quite a surprising result. Using a quadruple to summarize the results in 2014, (4, 6, 10, 0) indicates that there: were four perfect matches; six times when the #2 and #3 (or #1 and #4 – or both!) teams had their ranks reversed – which would not have changed how the teams were aligned for the opening games in the CFP; ten other times when all top four teams were correctly selected – albeit in a different order; and no occurrences where a team in the committee's top four tied for fourth place in any of the 32 rankings produced (by these models).

However, in 2015, the performance of these 32 models was quite different: (0, 0, 7, 3). The committee's top three teams were listed in all but one of the top four rankings that were generated when all 32 combinations were applied, but the other team came from the next four spots in the committee's final ranking. The team that the committee ranked as #4, Oklahoma, was ranked as the third or fourth best team ten times. The committee's fifth team, Iowa, who was undefeated until they lost the Big 10 championship game (to Michigan State – who probably earned the #3 position, in the committee's viewpoint, with that win), was (tied or) ranked as the third or fourth best team fourteen times, whereas twice defeated Stanford, who the committee ranked as #6, appeared as the third or fourth team (with some ties as well) eleven times. Finally, Ohio State, who the committee placed at #7, and who relinquished their place in the Big-10 conference championship game with their regular season their home loss to Michigan State, was ranked as the fourth best team, or tied for that rank, a total of five times (according to the 32 models proposed here).

Evaluating Convergence Alternatives

Given the somewhat inconsistent observed behavior of the 32 model combinations, with regards to the matching of the CFP committee's ranking in both 2014 and 2015, it seemed appropriate to expand beyond the one pass/single iteration approach which appeared to be sufficiently accurate regarding the matching of the CFP's decisions in 2014, but was much less so in 2015. Beginning with the aforementioned 'reasonable', initial ordering strategy, the rankings produced after each iteration were then used to create the groups for the next iteration, before performing new calculations for each of the models, until the rankings stabilized. (Occasionally, a few teams would oscillate indefinitely between two positions after each iteration, while all other teams' positions remain unchanged. For such 'unstable' teams, these oscillating positions could be averaged, though it may be unnecessary to do so in most cases.)

Most models required between 8 and 12 iterations before convergence occurred; the largest observed number was 22. One concern, however, was: how sensitive are the final rankings, produced by these models, to the initial ranking used? Similar results are produced when the ordering generated by the chosen initial ranking strategy is reversed, before the first iteration is made; however, any discrepancies whatsoever between those two rankings indicates that the initial ordering does have some impact on the final ranking that is generated (by any of the models, using this strategy). In an attempt to eliminate any initial ranking bias, one million random initial rankings were produced (for each model), and the average rank for each team was determined.

The results when applying this convergent strategy, starting with the (currently unspecified) 'reasonable' ordering, were (1, 2, 5, 1) in 2014, and (0, 0, 13, 0) in 2015. For the reversed version of those same reasonable rankings, 2014 yielded (1, 0, 4, 0) and 2015 (0, 0, 9, 2). When applying random ranking, convergence produced (0, 9, 0) in 2014, and (0, 0, 10) in 2015, where now it is virtually impossible for teams to have earned the exact same number of points for a model (over all one million random rankings), thereby eliminating ties in the final ranking (and the quadruple is therefore reduced to a triple).

Since none of these three, convergent approaches consistently matched the CFP committee's rankings in both 2014 and 2015 exactly, further study was required. However, before considering a somewhat different strategy (in the next section) to address this shortcoming, it seems prudent to first take a closer examination at the team rankings produced by these 32 combinations **outside** of the top four teams at this time.

It seemed somewhat egregious if a team not listed in the CFP committee's top 25 was awarded a rank less than 15 (as produced by any of the models described herein). For all 32 such models, where only one iteration was applied, none of them had such an occurrence for the 2015 data, and only two of them did in 2014 – when the 12-1 Marshall team was ranked 13th. This is fairly benign as compared to the situation when iterations are allowed to continue until the rankings have stabilized/converged.

When beginning with the teams, as ordered by the aforementioned reasonable ranking strategy, (which is utilized throughout this study), six of the 32 models had such violations with respect to the 2014 data, and five violations were noticed with the 2015 data (and only one model did so in both years). The worst case was produced by the EQ model, when using the 2014 data and the normal and reverse initial ordering (the EQ model was also the only model, relying on this strategy, with such a violation in both 2014 and 2015): Texas A&M (7-5) was ranked 9th, and Arkansas (6-6) was 10th by EQ. This 'anomaly' occurred primarily because in 2014, A&M defeated Arkansas, who had defeated Mississippi, who ended up ranked #3 (in both of these models) because the Rebels were the only team to defeat Alabama, who remained ranked at #1.

This specific situation is somewhat similar to the continual team rotation scenario alluded to much earlier in this article, where the points awarded may be too high, for wins over strong competition (teams in the most, or second most, selective group), and so teams with such wins tend to remain in the most selective groups as well. Overall, only 14 of the 32 models did not experience any violations in either year, using any of the four possible iteration strategies: one iteration alone, or, to iterate until convergence occurs – starting with either the regular, reverse or random ordering (where the latter utilizes one million random, initial orderings). Surprisingly, all four W orderings (WF, WL, WQ and WE) were without any violations, and seven out of the eight linear increase models (all but FL) avoided any such violations as well (which accounts for 10 of the 14 'clean models'). So, the intersection of the two model selection criterion yields WL as the most likely choice to avoid such inaccurate rankings (by the model). As it turns out, when using the reverse order initially, WL did correctly select the top four teams in 2014 and 2015 – with only the teams' specific ranks being different from those on the CFP committee's list.

Considering Margin of Victory

Up to now, the models described have only considered wins and losses, when attempting to order the teams in the set of the NCAA's division one football teams. In both prior years (2014 and 2015), the only undefeated teams were the champions of the Atlantic Athletic Conference (ACC). In 2014, Florida State was ranked #3 by the CFP committee, and in 2015, Clemson was deemed to be the top team. So, the obvious question is: what guided the CFP committee to these (at first glance) seemingly inconsistent rankings? (When no MOV is included in the aforementioned 32 models, and all four iteration strategies are considered, the models disagree somewhat with the CFP as well: the count for Florida State's rankings, in 2014, were #1 17 times, #2 60 times, 28 times as #3, 8 times being ranked #4, and 15 times ranked #5 or higher, whereas Clemson's counts, in 2015, were 2, 49, 57, 20, and 0. So, both teams appeared as #2 or #3 more often than the other ranks, but contrary to the CFP final ranking, Clemson was actually ranked lower, on average, than FSU.)

Considering both team's strength of schedule, utilizing the rankings produced by the Power Rating System (PRS) developed by Carroll et al [1], and considering only wins/losses, i.e. limiting the margin of victory to be at most one point, Florida State (in 2014) defeated teams who were ranked #14 (10-3) , #22 (9-3), #31 (7-5), #37 (6-5), #42 (7-5), #43 (7-5), #50 (6-6) and #51 (6-6) whereas Clemson (in 2015) had a few more high profile wins as they defeated the #9 (10-2) team as well as #19 (10-2), #20 (11-2), #38 (8-4), #45 (10-2) and #48 (7-5). Florida State also won seven games by less than seven points (in games against all of the opponents listed above – except #50), and most of those outcomes were very much still in doubt far into the fourth quarter of those contests; Clemson, on the other hand, only had three such close games.

Utilizing the PRS once again, Tables 2 and 3 depict how much difference the margin of victory (MOV) does make in the rankings produced by this recognized evaluative strategy (which incorporates strength of schedule equally with MOV). Florida State's rank, in 2014, was in the top four as long as the MOV was smaller than eight, then its position clearly dropped significantly with larger MOVs. However, in 2015, Clemson did not experience the same degradation as the MOV increased – staying between #3 and #5 regardless of the MOV selected.

The members of the CFP committee were probably also somewhat influenced by final scores, and not just who won – and who lost. A 20 point win, over a quality opponent, would probably increase the committee member's confidence concerning the victor's superiority more than a 1, 2, or 3 point win would. Therefore, to allow for MOV to be included into the 32 models already described here, whereby teams are awarded points, and eventually ranked – according to such final totals, a multiplier for each win/loss will be incorporated, via some transformation (of the actual MOV) that would also attempt to limit the influence of large wins. (During the BCS era, it was conjectured that some teams might have run up the score, against weaker competition, in an attempt to influence the quantitative methodologies that were in place then to rank teams, in an effort to improve a team's likelihood of being invited to the championship game.)

Margin of Victory	1	2	3	4	5	6	7	8	9	10	15	20	30	All
Alabama	2	3	3	3	2	2	2	2	2	2	2	2	2	1
Oregon	3	2	2	2	1	1	1	1	1	1	1	1	1	4
Florida State	1	1	1	1	3	3	4	5	5	5	7	11	16	18
Ohio State	4	4	4	4	4	4	3	3	3	3	4	4	6	6
Baylor	9	8	9	10	10	10	10	9	9	11	10	8	9	8
TCU	6	6	6	5	7	7	7	8	7	7	6	5	4	2
Mississippi State	10	9	8	8	8	8	8	6	6	6	5	6	7	7
Michigan State	13	12	13	14	13	14	14	15	15	14	13	12	11	11
Mississippi	8	7	7	7	5	5	5	4	4	4	3	3	3	3
Arizona	7	10	10	9	9	9	9	10	14	15	18	18	23	27
Kansas State	17	21	18	18	21	21	22	23	24	25	23	19	13	14
Georgia Tech	14	14	12	11	14	13	13	11	10	10	9	9	14	16

Table 2 – top 12 teams in the 2014 CFP standings, with ranking according to the PRS

Some previous suggestions have been to cap the MOV at a certain threshold (e.g. 20 points) and then possibly add on only small increments for the remaining points above that threshold. (The square root of the amount greater than 20 has been mentioned.) However, the application of this linear function, until the threshold is reached, implies that a 15 point win is therefore (somehow) 'worth' five times as much as a three point win. This difference, concerning these two outcomes' relative 'worthiness', seems to be too great, and so a different strategy was devised and implemented.

Margin of Victory	1	2	3	4	5	6	7	8	9	10	15	20	30	All
Clemson	3	3	3	3	4	5	5	5	5	5	4	4	4	4
Alabama	2	2	2	2	1	1	1	1	1	1	1	1	2	2
Michigan State	1	1	1	1	2	2	3	4	4	4	5	6	7	12
Oklahoma	4	5	5	5	5	4	4	3	3	3	3	3	1	1
Iowa	7	7	8	7	7	8	8	8	8	8	9	12	19	26
Stanford	6	6	6	8	8	7	7	7	7	7	6	5	5	5
Ohio State	5	4	4	4	3	3	2	2	2	2	2	2	3	3
Notre Dame	9	8	7	6	6	6	6	6	6	6	7	7	6	6
Florida State	19	17	16	15	15	16	15	13	12	12	10	8	8	8
North Carolina	20	19	19	19	18	18	16	16	15	14	11	11	9	9
TCU	12	11	9	9	9	9	9	10	10	10	14	15	13	13
Mississippi	15	14	12	11	10	10	11	11	11	11	15	13	14	7

Table 3 – top 12 teams in the 2015 CFP standings, with ranking according to the PRS

Rather than begin the MOV transformation after a certain threshold, whose determination might also be subject to debate, it made more sense to try and create a methodology that would cap the multiplier as well as provide reasonable relative ratio values for some small – and large – MOV values. The (base two) logarithm function was chosen because it does effectively reduce many large numbers to a small range of values as well as being coincidentally aligned somewhat with the common point values awarded for field goals and touchdowns. For instance, using the translation formula given in the top of the rightmost column in Table 4, winning by a field goal produces a multiplier of three while winning by a touchdown (and successful extra point try) yields an only slightly larger multiplier: four. In fact, any game whose outcome can still be impacted by essentially one 'big play', i.e. a contest whose MOV is seven points or less, will produce multipliers between two and four. Winning by three points is only worth 50% more than by a single point, even though the MOV is three times larger; a win of seven points is only worth 33% more than a three point win even though seven is more than twice the size (of three).

Winning by two touchdowns would be worth twice as much as winning by one touchdown, when using any sort of linear transformation (for scores less than the specified threshold), but with this (binary) logarithmically-based computation, the multiplier increases only by 20% (in this case). Roughly doubling the MOV again (to 31 points, which is slightly more than four touchdowns) again only marginally increases the multiplier, and so on.

One was added to the MOV so that a tie game would also be worth one half of a one point victory; in the National Football League (NFL), a loss was worth half a win in the standings before the single overtime period was added (to regular season games). The other plus one, which is added to the value produced by the $\log_2$ function, was chosen to avoid starting at a multiplier of one, for a one point win, since any other outcome multiplier, K, would then be K times more significant (though the current ratio, K/2, contributes half as much than if the plus one was omitted). This *a priori* chosen proportion seemed like it would perform better than **not** adding one – on an intuitive level. The result of including these two increments also seemed to generate a striking symmetry (as shown in Table 4) between integral values and frequent final score differences.

MOV	$\log_2(\text{MOV}+1)+1$
0	1
1	2
3	3
7	4
15	5
31	6
63	7

Table 4 – translation of (some) common MOV values (that produce integral multipliers)

The results after applying this particular MOV translation are as follows. (Rating/ranking ties are non-existent here, as they were with the random ranking approach, so triples can summarize each approach's performance.) In 2014, there was still strong, top four recognition (2, 17, 10), but less so in 2015: (0, 0, 10). For the convergent strategies, the results for 2014 were (2, 9, 4), (0, 4, 3), and (1, 1, 8) when teams started in: the aforementioned 'reasonable order'; in the reverse of that order; and with random ordering, respectively. For 2015, the performance was (0, 3, 4), (0, 2, 11), and (1, 2, 10).

Most Appropriate Models

With 32 possible model combinations, and four different, rank producing alternatives to go along with the decision to include (or not include) MOV, this yields a grand total of 256 different, generated rankings. When ignoring MOV, the overall performance of the 128 models was (6, 17, 19, 1) in 2014, and (0, 0, 39, 5) in 2015, so there was no model that perfectly matched the CFP committee's rankings in both 2014 and 2015. For the approach incorporating MOV, the results were somewhat improved: (5, 31, 24) in 2014 and (1, 6, 36) in 2015. Unfortunately, perfect matchings for both years did not materialize here either.

However, of the six perfect matches, three models correctly chose the top four teams in the other year. Even better, there were two models that ordered the top four teams so that the CFP matchups were identical to those produced by the CFP committee in those two years. These two models (NL and WL) utilized the convergent strategy, starting with the 'reasonable' team order, and neither of these models suffered any violations amongst the top 15 teams, in either year; both models are also based on what seems to be the most consistent point value choice (linear increases) and the two grouping strategies in use are both variations of Zipf's Law.

Table 5 lists how these two models ranked the top four teams, and it also includes the Spearman Correlation Coefficients (SCCs) which provide a measurement for how closely these models matched the top 25 teams listed in the final CFP ranking. Since teams #1 and #4 play are scheduled to play each other, as are teams #2 and #3, one can easily observe that NL and WL both have the same matchups as produced by the CFP committee, though the actual top four rankings are somewhat different (when Oregon faced FSU in 2014, as did Alabama and OSU, while Alabama and MSU played in 2015, as did Clemson and Oklahoma).

Four teams were seven or more places away from their CFP ranking in this 2014 NL ranking; #11 Kansas State was the largest disagreement (in these four), having been ranked #24 by this model. Four teams were also in this category in 2015 (for NL), with #19 LSU finding itself to be the furthest outlier – being ranked eight places higher. Regarding WL, it had one team off by eight, nine and ten positions, respectively, in 2014, but there were five such teams in 2015, three of them being off by eight, one by seven and the other one by ten. (Houston was #18 according to the CFP committee, and ranked by WL as #8; as it turns out, Houston did defeat the #9 CFP team, Florida State, in their post season, 2015 bowl game.)

With regards to these particular convergent models (NL and WL, with the initial, 'reasonable' order), the only teams in the committee's top ten, which were more than five positions from where the committee ranked them were: #5 Baylor in 2014, which was #15 according to the NL model, and #10 North Carolina in 2015 – ranked #17 by WL.

2014	NL	WL
1 – Alabama	3	2
2 – Oregon	1	1
3 – Florida State	4	4
4 – Ohio State	2	3
SCC	0.7869	0.8246
2015	NL	WL
1 – Clemson	2	2
2 – Alabama	1	1
3 – Michigan State	4	4
4 – Oklahoma	3	3
SCC	0.8377	0.8292

Table 5 – Top four team placements, according to the top two matching models

The random ranking approaches do have a little more statistical merit than the two models elaborated above, since they are the product of one million random orderings. The average SCC for the 32 models in 2014, with no MOV consideration, was 0.6543, with ten models below 0.6 and three above 0.8 (high of 0.8250; the WF model). In 2015, the SCC value was 0.7013, with one above 0.8 (0.8019; NF) and one below 0.6 (0.5777; HE). When incorporating the MOV translation, the SCC value for 2014 increased to 0.7590, and only three models were below 0.6, while ten were above 0.8 (high of 0.9058; GF); however, in 2015, this value shrank a little, down to 0.6881, with nine models below 0.6, and nine above 0.8 (high of 0.9304; HL).

The best average SCC (with initially, random team placements) for both years was 0.8554 (HL); this model correctly predicted three of the top four teams in both years, and the point value for the team left out was only 0.52 too low in 2014, and 1.56 below the incorrect #4 team's point value in 2015. Modifying one or more of the constants in the MOV translation could rectify this shortcoming, but it could also negatively influence NL, where the #4 team was only 0.05 above the #5 team. (Average point totals for these 'bubble teams' range in value between 70 and 90.)

A sensitivity evaluation was also performed on two different, one million random initial orderings, for one of the models (VL), and 96 of the 128 teams had final rankings that were identical down to the hundredth's digit; the other 32 teams were off by just one hundredth of a point (when the total, average point totals per teams were output as XXX.HH).

A Few Reflections

There are certainly many other possible grouping strategies and/or point value assignments which could be considered (as well as other MOV translations – and/or different constants in the current MOV translation). This study has illustrated that there is some relationship between the approaches incorporated into all of these models and the final rankings that have been determined by the CFP committee in their first two years. Even though a model has not been discovered that exactly produces the same top four teams as the CFP committee, other strategies (or approaches) are still possible.

For instance, perhaps the totals from two or more of these models could be averaged, or, after generating the ranking produced by one millions random initial rankings, one final iteration (applying the same model) might produce the desired outcome. Further investigation into these models, especially given more data over the next few years, could help to narrow down the choices for a viable, objective methodology that could imitate how the CFP committee 'thinks'.

Conclusion

It may be impossible to discover a totally objective approach that will always produce the exact same rankings as they are currently being subjectively generated. However, it seems clear from all of the models described herein, that there is a strong level of agreement between many of these models, and the rankings announced by the CFP committee. In most cases, there is agreement on three – and sometimes even four – of who the committee thinks are the top four teams that year. It will be interesting to see if the NL and WL models continue to agree with the committee's choices, and perhaps other researchers will contribute new (or modified versions of these) models for future evaluation/consideration. Given that it is possible and/or likely that some amount of bias could enter into any decision where humans are involved, it would be nice if such decisions could be validated – by a method that produces the same result, where this method would exclude the subjective component of the human decision making process.

References

[1] Carroll, Bob, Palmer, Pete, and Thorn, Jim (1988), *The Hidden Game of Football*, Warner Books.