Objectively Modelling the College Football Playoff Committee's Selections

John A. Trono
Saint Michael's College
Colchester, VT
Tech report: SMC-2019-CS-001
(Presented at the 7th MathSport International Conference, July, 2019)

Abstract

At the end of the NCAA college football season, four teams are selected (by said committee) so that those teams can then compete in the single elimination, College Football Playoff (CFP) which crowns that sport's national champion. Thirty two models have been evaluated – over the first five years of the recently created CFP, with regards to matching how the CFP committee has ranked its top 25 teams – by varying the *victory reward* parameter (four choices) and the *group-size* parameter (eight choices) when employing one possible strategy to assess each team's success that year. Those 32 models only consider wins and losses, whereas a slightly modified set of 32 models includes margin of victory (MOV) as well. One of those models – both with, and without, MOV – has correctly predicted (the same) 18 of the 20, top four teams that the CFP committee has chosen the past five years, and another ten models (seven using MOV) have matched 17 out of 20. These various models, and a deeper analysis of them, will be presented here.

Introduction

In most sports, the outcomes of competition – between teams or individuals – typically leads to the determination of a champion, e.g. the World Cup, the French Open, the (British) Open, the men's – and women's – National Collegiate Athletic Association's (NCAA) college basketball tournaments (also referred to as *March Madness*), etc. However, the NCAA football season, at least until 1998, relied solely upon the opinions of many sportswriters (the Associated Press – AP – poll), or many college football coaches (the United Press International – UPI – poll), to crown its National Champion (NC) in that sport. Prior to 1990, only occasionally did those polls disagree as to who they thought was the *best* team that year (and only once from 1975 to 1989).

Background

After the conclusion of the regular season, teams who had distinguished themselves during that season would be invited to play in one of several postseason 'bowl games'. Top teams from the (perceived) strongest football conferences would meet in the Rose, Orange, Sugar and Cotton Bowls on New Year's Day. The final polls (after 1967 for the AP poll, and after 1973 for the UPI poll) would take into account all such postseason outcomes, and these poll results were initially recognized as accurate depictions of who the top ten (or twenty) teams in the country had been that year. (There would always be some arguments regarding certain teams being ranked too high – or low – in those polls, but typically *not* about who was the #1 team. There also appears to have been a significant increase in interest, regarding college sports in the 1970s, so when the polls differed on who they thought was #1 – in 1978 – a movement began to consider a different approach to who was #1.)

With over 100 teams to choose from, and roughly five to ten bowl games being played between 1935 and 1970, these contests were fairly selective, and matched up teams who had proven themselves on the field, typically only losing two or fewer games (over their ten game season) – or at least that was the situation until 1970. Over the years, an additional bowl game (or two) has been added occasionally, somewhat diluting the quality of the teams being invited (in the newly added, less prestigious bowl games): there were 14 bowl games in 1980, a total of 18 in 1990, 25 in 2000, and 35 by 2010. A saturation point has now been reached where almost all teams who win at least half of their twelve games are eligible to be invited to one of the roughly forty, extant bowl games – with only approximately 130 teams to choose from. (In 2015, three teams with 5-7 records were invited – to three different bowl games – because all the teams with records at or above 0.500 had already been invited, or were prohibited from playing in a bowl game due to NCAA sanctions.)

Before 1990, there was usually one year each decade where some controversy would occur when both polls disagreed as to who they voted as that year's #1 team. However, there were four years in the 1990s (before 1998 – when the Bowl Championship Series (BCS) first began) where such disagreements eventually produced a methodology so that the NC could be decided on the field (and not by voters' opinions). Before 1998, the primary reason for the strong likelihood of such controversial recognition of the NC was linked to which conferences had become affiliated with which major bowl games. When only two undefeated teams remained, at the end of the 1991 season, it would've pleased everyone to have them play each

other. Unfortunately, the Pacific-10 conference champion, Washington, was committed to play against the Big-10 conference champion in the Rose Bowl instead; the other undefeated team, the conference-independent Miami (of Florida) team, was invited to play in the Orange bowl.

Both of these undefeated teams won their respective bowl games, and the AP and UPI polls ranked them as the #1 and #2 teams, albeit in different order. When the same thing happened again in 1994 (though both polls ranked the two undefeated teams in the same order that year, with some controversy still extant) – and in 1990, with somewhat different, specific circumstances – a system was eventually devised by the NCAA so that the best two teams would meet in a postseason contest to avoid this designation of *split NCs*. (Sadly, *split NCs* occurred again in 1997, the year before the BCS began.)

Even though the BCS did prevent the likelihood of a split NC from occurring, roughly half of the BCS years (1998-2013) created a different topic of controversy: who were the best two teams? The BSC ranking formula utilized several deterministic strategies: results from two polls, and some number of computer-based, evaluation systems/algorithms. Unfortunately, when the results produced didn't satisfy the NCAA, adjustments to the system were made: modifying which computer systems were included; bonuses were added for wins over teams in the BCS's top ten; the margin of victory (MOV) was to be excluded by all the computer-based systems, etc.

Finally, it was decided that the number of teams vying to be the NC would be increased to include four teams (though some experts believe that eight would be a better number), and a committee was established to decide which teams would be invited to the College Football Playoff (CFP). The CFP committee, following guidelines established to aid them during their deliberations, would make weekly, preliminary rankings after more than half of the season had been played, and the committee's composition would change some from year to year. (A more detailed history of the first two years of the committee's voting can be found in Trono, 2016; even with the expansion to four teams, there was still some disagreement regarding the fourth/last team selected in 2014, who also ended up as the NC!)

The Models

The basic principle for constructing the models to assess each team's success in a given year follows a primarily discrete, rather than continuous, approach. The first step is to separate the teams into distinct groups, whose sizes became smaller as the

teams in them have demonstrated more success that year. There are eight different *group-size* strategies for creating these groups, and each win over a team in a more selective (and smaller) group is rewarded more points than a win over a team that resided in a larger, less successful group. Four different *victory reward* functions were also studied, and will be described shortly. However, since every team doesn't necessarily play the same number of games (in a given year), and some teams have occasionally *padded* their schedule with *weak* opponents, it also made sense to provide no reward for defeating these *weak* teams. Therefore, all four functions assign zero points for the least selective (and largest) group of teams.

The most straightforward victory reward function is linear (L: 0, 1, 2, 3 …), and a win in each more selective group is worth one more point than the preceding group. For the exponential (E) function, a win in the next group is worth twice as many points as the preceding one (0, 1, 2, 4, 8 …). The quadratic function (Q) increases by one the number of points that separate how much a win in one group is worth over the preceding group (0, 1, 3, 6, 10 …); lastly, the Fibonacci function (F) uses the values (after excluding the initial two ones that appear) in the Fibonacci sequence: 0, 2, 3, 5, 8 and so on.

The eight different group-size strategies also follow several, similar guidelines – for increasing the exclusivity of each group, as each subsequent set shrinks in size. The exponential strategy (E) places 50% of the teams in the first group, and each group thereafter is half the size of its predecessor. (However, this process stops when the last two groups both contain roughly one sixteenth of all the teams.) Using consecutive fractions, in a manner motivated by Zipf's Law, is the basis for the N strategy, where group sizes are 1/4, 1/5, 1/6, 1/7, 1/8 and almost 1/9 (where N was chosen because of the final group size – one ninth). The W strategy starts the first group size at 1/3, then 1/4, 1/5, 1/6 and the last, most exclusive, group is then exactly 1/20 (one tWentieth). Using only odd integers in the denominator, the O strategy group sizes are: 1/3, 1/5, 1/7, 1/9, 1/11, 1/13 and then roughly 1/23 in the last group.

Starting with forty percent, the F strategy reduces the size of the next group size by 10% (yielding groups of 40%, 30%, 20% and 10%). The V strategy (V is the Roman numeral for 5) is similar, but includes only 5% reductions for the first and final group determinations, with all others being 10%, i.e. 35%, 30%, 20%, 10% and 5%. The H strategy partitions the teams into two halves, where half of the first half receives zero points, and the other half (25% overall) receives the smallest victory reward. The other, more exclusive half is divided into four groups – each 5% smaller than

the preceding one: 20, 15%, 10% and 5%. The G strategy attempts to reduce each group by the golden ratio; to achieve this goal, and have a reasonably small final group size, the first group contains 37.5% of all the teams. The next group is 70% of that size, (i.e. 26.25% of the overall teams), and each subsequent group is reduced to 61.8% (the reciprocal of the quantity $(1+\sqrt{5})/2$, representing the expected increase between terms in the Fibonacci sequence, after enumerating a reasonable number of terms) of its predecessor's size.

Monte Carlo Approach

As one might surmise, the initial placement of teams within any grouping, will inherently impact how many points are awarded to each team for each of their wins. (More details can be found in Trono, 2016.) To reduce any bias – from any possible strategy to determine an appropriate, initial team order – one million random orderings are generated. For each ordering, every team's season is evaluated, given the underlying victory reward and group-size selections, and the new ordered list, sorted by the total points accumulated, was then fed back into the same model until every team's position has stabilized in that list so that every team remains as a member of the same group (and would thereby produce the same identical point total per team on any subsequent iteration of this awarding of points). Each team's point totals were then averaged, across all random orderings, producing the final ranking – with the four teams accumulating the highest average point totals being those that the model would choose to compete for the NC.

In 2014, nine of the 32 models (excluding MOV) matched all four teams chosen by the CFP committee, albeit in a slight different ordering of teams (from the committee's choices). Teams #1 and #4 play each other, in one CFP semifinal contest – as do the teams ranked #2 and #3 by the CFP committee in the other semifinal, so the order produced by the model can generate the same matchups as the committee if teams #2 and #3, and/or teams #1 and #4, were reversed in the ranking produced by any model. All nine of the aforementioned models did have the #2 and #3 teams in the opposite order of the committee, so these models would've established the same two semifinals as were chosen by the committee.

In 2015, ten models chose the same four teams as the committee, but there were no models that matched which teams should've been playing against each other in either of those two semifinal games, as established by the committee – whose #1 team was ranked as the #4 team by two models, and who was also listed as the #2 or #3 team fifteen times each (across the 32 models). This led to an investigation to consider

including MOV into the 32 models to possibly correct this shortcoming (Trono, 2016). Four models (not including MOV) did match the committee's top four team selections for the first two years: WF, WL, FF, and FL. (The victory reward value designation appears second in these two letter, model abbreviations.)

The details about why the MOV formula in Table 1 was chosen appear elsewhere (Trono, 2016). Essentially, the resulting correspondences between the smaller, integral MOV values, and several frequent, final score differences (in Table 1), along with the inherent properties of the (base two) logarithmic function (it doesn't significantly reward teams who *ran up the score* against weaker competition), were the deciding factors in choosing this particular MOV formula.

For each contest, the quantity determined by this formula is used as a multiplier of the victory reward value, coinciding with the group (for that model) that the losing team is in, and those quantities are all summed for each team. When a team loses, a negative value is produced – both with and without MOV – and that quantity also contributes to the team's total. Losing to a strong team is akin to defeating a weak team, so the victory reward value is a small, negative value, whereas losing to a weak team will coincide with a larger victory reward value (similar to defeating a strong team, though a negative value is used in this situation as well), thereby inflicting a more severe penalty. In a sense, the victory reward values are determined in *opposite directions*, when using them for wins and losses; however, +1 is added to each victory reward value selected for a loss (before that value changes from positive to negative), so that there is a small penalty even when losing to a team in the most selective group.

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

Table 1 – Some common MOV values (that produce integral multipliers)

As it turns out, there were also nine models, when the MOV is included in the ranking process, that selected the same four top teams as the CFP committee did in

2014: one model (WE) had the four teams ranked in the same order as the committee; another model (HE) would've scheduled the same two contests (though that model's ranking was quite different from the nine aforementioned models that exclude MOV, as the CFP top four were ranked in the order 2, 1, 4, 3 by HE – when ignoring MOV); and the other seven models would've set up two different contests (from those that the CFP decided upon).

There were thirteen models that matched the committee's choices in 2015 when MOV was used: one matched exactly (EL); two had the same scheduled games (NL and VL; both also ordered the top four 2, 1, 4, 3); and the other ten established different matchups. There were just three models (WL, FL, and FE) that included MOV and matched the eight teams selected by the committee in 2014 and 2015.

In 2016, five models that ignore MOV matched the committee's top four teams, with NL and WL matching their ranking exactly, and VL, HL and GL having #2 and #3 in the wrong order (but still agreeing with the committee regarding which teams would compete against each other). All five of those also matched the top four when MOV is included along two more models (WF and VF), however, all seven of these would've established different pairings for the two semifinal contests. (WL is the only model to match all twelve teams that were selected by the CFP committee during the first three years of this playoff system – both with and without MOV.) (The undefeated Western Michigan team was only ranked as the #15 team by the committee in 2016, though this team was mostly ranked from #9 to #13 by the 64 models described herein.)

In 2017, another undefeated team (from a different *mid-major* conference than Western Michigan) was also excluded from the CFP, ending up as the #12 team in the final committee vote, and all but six models (without MOV) had that team ranked as either the #2, #3, or #4 team that year. All four models relying on the group-size parameter 'N' ranked the University of Central Florida (UCF) between #9 and #12, whereas OL and HF ranked them just outside the top four (as #5). In a similar fashion, the four 'N' models ranked UCF from #10 to #13 – even when MOV was included – while ten more models ranked UCF as #4, and the remaining 18 models listed UCF as #5. (UCF did earn two quality wins over conference rival Memphis (10-2) and one over the 9-2 South Florida squad. And even though the eight 'N' models didn't rank UCF in the top four, the four models including MOV replaced UCF with Notre Dame, which was ranked as the #14 team by the committee, while three 'N' models that do ignore MOV chose USC, and the other such model placed

Wisconsin in the top four, where the committee ranked them as the #6 and #8 teams, respectively, that year.)

In 2018, a very strong Georgia team, which played a very competitive schedule, was highly regarded by all 32 models even though they lost twice (to top teams): to the undefeated #1 CFP selection (Alabama) in the Southeastern Conference (SEC) championship game (losing by only 7 points) as well as incurring a road loss to a strong LSU team (who finished with a 9-3 record, and who also is a member of the SEC). All 32 models, with and without MOV, had the Bulldogs ranked between #2 and #4 this year. UCF was undefeated in 2018 as well, and were ranked as the #8 team by the CFP committee – up four places, from 2017 (perhaps due to their increased credibility after their bowl win over Auburn in 2017) – though only eight non-MOV models ranked them as high as #5 while a dozen MOV models ranked them as #5. One win over Cincinnati (10-2), and another over Temple (8-4), were the only *resume building highlights* this year, and strangely enough, the objective power ranking system also rated UCF higher in 2017 (than 2018), like the models did, contradicting the voting pattern produced by the committee. According to the power ratings that were calculated, UCF was ranked #5 (101.045) in 2017, when excluding MOV, and #8 in 2018 (100.845), whereas when including MOV, UCF was #13 in 2017 (120.461), and #17 in 2018 (115.694).

	9	10	11	12	13	14	15	16	17	18
Zero	1	2	2	4	4	4	6	5	3	1
MOV	0	0	0	2	3	6	7	6	7	1

Table 2 – Number of top four CFP teams as selected by the models from 2014-18 (Zero indicating that Margin of Victory – MOV – was excluded).

Table 2 illustrates how effective all of the models were in matching the CFP committee's selections. WL is the only model to match 18 of the 20 teams selected by the CFP committee over the first five years of its implementation. One can also observe, that on an overall basis, the models including MOV have slightly outperformed those that exclude MOV, averaging 15.15625 correctly matched teams versus 13.9375 teams. (WF and VL are the only two models that matched 17 correctly – both with and without MOV. HL is the only other model to match 17, excluding MOV, and the five other models with 17 correct, where MOV is included, are NL, VF, FL, FE, and GL.) Six teams were correctly chosen to be in the top four

by all 64 models (32 with, and 32 without MOV): (ranked #1 by the CFP committee) Alabama in 2014; (#3) Michigan State in 2015; (#3) Ohio State in 2016; (#2) Georgia in 2017; (#1) Alabama and (#2) Clemson in 2018. One more team was always ranked #1-#4 by the models that exclude MOV (#1 Clemson in 2015), and three more appeared in the top four of all models including MOV (#2 Oregon, in 2014, #1 Alabama, in 2016, and #1 Clemson, in 2017), so in total, there were seven teams that matched the committee's top four in all 32 models that ignore MOV, and nine teams that did likewise in the 32 MOV-based models.

And the Winner is …

A careful observer might have noticed that the most accurate models mentioned so far have overwhelmingly utilized the linear victory reward value (L). In fact, 14 of the 24 models, without MOV, that matched the top four, over these first five years of the CFP, employed the linear victory reward approach, with six using F, and two each for E and Q; likewise when including MOV, 14 of 29 used L, seven used F, there were five E's, and three Q's. When only considering perfect matches, and inconsequential rearrangements in the top four order, the linear victory reward strategy produced three of the four perfect matches, and nine of the fifteen that were inconsequential. (F produced four that were inconsequential, and E the other two in this category as well as the remaining perfect match.)

Even though matching the top four is the primary objective of the models being described herein, it would be relevant to also consider how well the models matched the final top 25 teams the committee produced. For this analysis, the Spearman Correlation Coefficients (SCC) were calculated for all models, thereby determining how well the models matched both the top four as well as the top 25 teams for the first five years of the CFP.

Top 4 Matched	Model No MOV	SCC Value	Top 4 Matched	Model With MOV	SCC Value
18	WL	0.540	16	HL	0.580
17	HL	0.420	17	VL	0.480
15	NL	0.340	17	NL	0.445
17	VL	0.340	15	WE	0.380
17	WF	0.280	15	WF	0.320

Table 3 – SCC values for top five models when matching committee's top 4 teams.

Top 4 Matched	Model No MOV	SCC Value	Top 4 Matched	Model With MOV	SCC Value
15	NL	0.7659	16	HL	0.7854
17	HL	0.7412	17	NL	0.7669
17	WF	0.7132	18	WL	0.7091
18	WL	0.7091	17	FL	0.7058
17	VL	0.6824	17	FE	0.7048

Table 4 – SCC values for top five models when matching committee's top 25 teams.

Upon examining the information contained in Tables 3 and 4, one notices that the same five models produced the closest matching of the committee's top four, and top 25, over the past five years, when ignoring MOV – albeit in different order, with these two measures. (Four out of five are also linear models.) Since WL was the only such model with 18 out of 20 correct teams chosen, it is not surprising to see them with the highest SCC value, when only considering the top four teams.

It was somewhat surprising to see that the WL models have exactly the same SCC value for the top 25 teams both with and without MOV. HL has a slightly higher SCC value for the top 25 teams, with regards to both types of models, however, WL's values aren't that much lower, and, it did match the committee's selections once (or twice) more than HL succeeded in accomplishing. (WL's absence from the MOV portion of Table 3, even though it too did match 18 out of 20 teams, is primarily due to having the #4 team in 2017 be ranked as #8 by this model as well as the #1 team in 2015 being ranked as #4 by WL. On the other hand, HL's four misses, when including MOV, occurred by ranking all four omitted teams – excluding 2016, when it matched all four – as #5, which helps to explain that relatively high SCC value in the MOV portion of Table 4.) With regards to the top four teams selected by the committee, from 2014-2018, the WL model (without MOV) had ten teams matching the committee's positions exactly, while the HL model only matched nine, and when including MOV, the WL model matched nine, and the HL model only five. (See specific rankings for WL in the Appendix.)

Conclusion

The WL model has matched 18 of the 20 teams selected by the CFP committee, from 2014 to 2018, and when ignoring MOV, it also attained the second highest SCC value for the top four teams, and the sixth best overall match when considering the top 25 teams (as produced by this committee) as well as correctly placing ten of the

twenty teams in the first five years of the CFP into the same position the committee placed those teams. In the WL model, all teams reside in one of five groups, with the least worthy one third (of the teams) in the bottom group, and the next three groups (of decreasing size, and, increasing success that year) holding 25%, 20% and 16.6% of the teams, while the final, and most exclusive group, contains the top 5% of all teams being considered. Wins over teams in the largest group aren't rewarded, but 1, 2 and 3 points respectively are awarded for wins over teams in the next three smaller groups, and 4 points are earned for a win over a team in the top group. (Losses to teams in that decreasing set of groups would accrue penalties of -5,-4,-3, -2 and -1, respectively.) This non-MOV model matched the committee's selections exactly in 2016, only had #2 and #3 (inconsequentially) interchanged in 2014, matched the top 3 in 2017 (with UCF as #4), and incorrectly included Georgia as #4 in 2018 while also considering eventual champion Clemson as the #1 team, in contrast to the committee's ranking them as the #2 team behind Alabama, who WL ranked as #2, with Notre Dame correctly ranked as the #3 team. In 2015, WL ordered the CFP #2 through #4 teams as #1 through #3 while ranking the committee's #1 choice as #4.

References

Trono J. (2016). Is it Possible to Objectively Generate the Rankings Produced by the College Football Playoff Committee? Saint Michael's College Tech Report (SMC-2016-CS-001).
http://academics.smcvt.edu/jtrono/Papers/CFP_And_Partitions.PDF, accessed 5/14/2019.

Appendix

The table on the next page indicates how the CFP committee's top four teams would be ranked, according to the WL models – both with, and without, MOV. It also contains the top five teams, according to both models, so that the actual, average point totals accumulated (over the one million Monte Carlo, initial orderings) can be compared. Each team's rank, as determined by the committee, is listed in the leftmost column, and you can see which teams were in the exactly correct positions: 10 when ignoring MOV, and 9 with MOV. (In the non-MOV model, Oregon and Florida State would still have played each other, in 2014, since their ranks were switched inconsequentially.) Both models agreed correctly with Alabama being #1 in 2014 and in 2015, and they both had the top 3 teams correct in 2017; when including MOV, Oregon was correctly ranked #2, in 2014, Michigan State #3, in

2015, and the top two teams were also correct in 2018, and – when ignoring MOV – Ohio State matched (as the #4 team) in 2014, all top 4 teams were matched exactly in 2016, and Notre Dame was correctly ranked #3 in 2018.

	2014			WL		WL
CFP	Team	W-L	noMOV	Rank	MOV	Rank
1	Alabama	12-1	23.03	1	106.26	1
2	Oregon	12-1	16.10	3	92.30	2
3	Florida State	13-0	21.16	2	84.64	4
4	Ohio State	12-1	14.64	4	73.30	3
9	Mississippi	9-3	13.28	5	67.41	5
	2015					
1	Clemson	13-0	15.88	4	77.06	4
2	Alabama	12-1	21.01	1	104.72	1
3	Michigan State	12-1	19.64	2	77.45	3
4	Oklahoma	11-1	16.59	3	79.45	2
7	Ohio State	11-1	13.93	5	65.17	5
	2016					
1	Alabama	13-0	25.76	1	144.77	1
2	Clemson	12-1	18.01	2	78.01	4
3	Ohio State	11-1	16.50	3	90.96	2
4	Washington	12-1	15.07	4	88.54	3
5	Penn State	11-2	12.25	5	50.41	6
6	Michigan	10-2	9.64	6	62.06	5
	2017					
1	Clemson	12-1	19.99	1	111.96	1
2	Oklahoma	12-1	17.51	2	91.49	2
3	Georgia	12-1	16.29	3	83.69	3
4	Alabama	11-1	13.25	6	61.77	8
6	Wisconsin	12-1	11.40	9	71.22	5
8	Southern Cal	11-2	14.17	5	45.94	12
12	Central Florida	12-0	14.96	4	72.27	4
	2018					
1	Alabama	13-0	24.93	2	124.41	1
2	Clemson	13-0	26.45	1	115.23	2
3	Notre Dame	12-0	20.16	3	90.19	4
4	Oklahoma	12-1	10.57	6	62.79	5
5	Georgia	11-2	19.95	4	93.51	3
6	Ohio State	12-1	11.27	5	36.14	9