ISSN: 2754-6705

Journal of Mathematical & Computer Applications

Research Article

Open Access

The Selection of the Competitors for the NCAA Football Championship: A Retrospective

John A Trono

Professor of Computer Science, Saint Michael's College, One Winooski Park, Colchester, VT, 05439, USA

ABSTRACT

A collection of various team rating, and ranking, strategies are described as well as some hybrid combinations of these strategies. How closely these approaches have matched the consensus regarding who the top college football teams have been, since the inception of having teams compete on the field for that sport's national championship, is examined along with an evaluation of how well some of these methodologies have matched the current committee who decides which teams will be invited to this competition.

***Corresponding author**

John A Trono, Professor of Computer Science, Saint Michael's College, One Winooski Park, Colchester, VT, 05439, USA.

Received: December 13, 2025; **Accepted:** December 17, 2025; **Published:** December 26, 2025

Introduction

Professional and collegiate sports leagues alike crown their champion at the conclusion of their respective, postseason playoffs/tournaments. However, before 1998, the designated National Collegiate Athletic Association (NCAA) football champion was sometimes controversial because no such methodology existed: two human polls were relied upon to identify the top team each year, and when those polls were in disagreement, that led to no one team recognized as being the best. This situation occurred once or twice a decade, before 1970. However, when the frequency of such occurrences increased, this led to the creation of the Bowl Championship Series (BCS, in 1998), and then the College Football Playoff (CFP, in 2014), both of which were proposed to eliminate that situation.

Prior to the BCS, most major conference champions were typically committed to play in specific, postseason bowl games – which precluded many matchups between who the two polls perceived as the top two teams; these contests could've resolved some arguments in certain years (1991, 1994 and 1997) when only two undefeated teams, from the top conferences, remained. (Thankfully, some years, back then, did allow for the meeting of the polls' #1 and #2 teams in a bowl game, specifically: 1982, 1986, 1987, 1992 and 1995.)

This paper will present many different strategies to either: attempt to objectively evaluate every team's performance in a given year, or, attempt to match which teams the CFP committee ranked high enough to be invited for that year's postseason tournament. Some novel combinations of these strategies will also be included herein, and a previous methodology will be expanded to consider the currently invited set of - who said committee decides are - the top twelve teams instead of just the top four teams that said methodology was designed to predict.

BCS Background and Postseason Predictions

As the BCS strategy was initially being devised (in 1992), a group began to deliberate upon how the top two teams would be selected (to compete in a championship game); it appears that they wanted to have a more objective way to resolve the situation when the two polls disagreed as to who the best two teams had been that year. Therefore, the resulting formula weighed the two polls' contributions quite heavily when generating that final ranking/ordering; that initial formula also included three computer system's rankings as well as a strength of schedule (SOS) component, and, a penalty for each team's loss. The next year, the number of computer systems was increased to eight, and over time, the formula continued to be adjusted. A quality win component/bonus, over teams in the BCS top ten, was added to the formula, that determines a team's BCS ranking, and then, after the 2001 season, it was deemed that the computer systems weighted wins by large margins of victory (MOV) too heavily, and so in 2002, the computer systems were to ignore MOV (restricting it to be at most one point); one system's creator would not comply and was therefore excluded. In 2004, another system chose to no longer be included (so the number of remaining computer-based systems was six) and the quality win and SOS components were also removed. (All the remaining computer systems rankings were averaged after – just one of – the lowest, and highest, ranking values were excluded.)

There is no objective definition that everyone would agree to, regarding how to determine who the best teams are: is it the two teams who have accomplished the most (which some entity would need to define), i.e. which teams have had the best season, or, should the teams that would be favored (via some rating system) against all other teams, on a neutral site, be considering as the best? Some articles on this topic have been written [1,14] and another one concluded with "This article demonstrates that using the margin of victory results in a clear advantage in predicting future games." and, "I would not suggest that the BCS standings use a computer rating system exclusively – after all, they are interested in accomplishments as much as raw skill" [2].

There are hundreds of possible candidates for which computer-based systems could be included in that specific, BCS formula component (two examples are described in [12], and [13]; the former was included by the BCS starting in 2001, and the latter in 1999); this article will rely heavily on the (PRS) Power Rating System (as described in [11]; please see the Appendix for a brief description of the PRS). A different article evaluated PRS and many other similar systems, over 26 years of NCAA postseason contests and included the results of these system's predictions [3]. Using the full MOV, the PRS made the highest number of correct predictions (363, and 209 incorrect ones – with 5 bowl games ending in a tie, between 1983 and 2008, inclusively); restricting the MOV to 17, 7 and then 1, the PRS did progressively worse: 344-228, 341-231 and 328-244, respectively. The Las Vegas line, for that same timeline, was 350-206; 16 games were determined to be either a toss-up, or no specific line was established – due to one or more injuries (to key players). In a shorter time window, 42 online systems were evaluated over a seven year period (2002-2008), and the PRS was tied for seventh, with 129 correct predictions, and the Las Vegas (LV) line was one behind the top system with 135 correct predictions (out of 214 games); the LV line did much better than most systems over the last two years in that study, rising up to almost *take the lead*, so perhaps, over time, more data analytics were being considered when that line was being established? (The lowest total for any of the 42 computer systems in that study was 113 correct predictions, during that seven-year period, where between 28 and 34 bowl games were played each year.) Over the past decade or so, many more collegiate players are now deciding not to play in such postseason, college bowl games, in an attempt to avoid possible injuries, before these players became professionals the following year; this makes it more difficult to validate which systems make more accurate postseason predictions (when key players that contributed to a team's rating are absent).

The Rewards System

In 2002, the BCS began enforcing that all of the included computer systems must ignore MOV, and shortly thereafter, an interesting paper supporting this idea was published [4]. This article provided a model (Penalized Maximum Likelihood) that relied only on win/loss data, and the model yielded the smallest average discrepancy between where teams finished in the polls, after the regular season ended, and where the models in the BCS ranked those same teams. If such trust is placed with those voting in those two polls, i.e. relying upon those voter's expertise to make such an evaluative set of ranking decisions, then perhaps trying to objectively model both polls' behavior could produce beneficial results.

One observation, concerning how teams are usually ordered in these polls, led to a new model : if teams from the top conferences had identical records, then typically, the team with the most impressive win would be ranked above *similar* teams (with identical records). This recognition seemed to pervade the actual rankings in many years of both polls; some top teams had played quite a few weak teams, but if they had a win over one of the top teams, this earned them a high ranking in the polls. In the Rewards ranking approach, the PRS that ignores MOV (aka NO_MOV) was provided with all games that year (where all non-Division I opponents were lumped into one name: NON_DIV1A), and the NO_MOV ratings were used to determine a team rating for all teams (creating an ordering/ranking of all teams-via that final, Rewards rating) [5].

Each team's wins are ordered from the largest to the smallest NO_MOV rating value (with the highest overall NO_MOV rating being normalized to essentially one, and the lowest one to essentially zero), and a multiplicative weight was assigned to each of those game's ratings (for that team); when traversing the list, which is in descending ordered, the next weight is half of the previous weight, starting with one for the highest defeated opponent's rating. Then, a weighted average for all those victories was calculated and essentially multiplied by the number of wins by that team. (The normalized NO_MOV rating for each loss was also then subtracted from this total as well.) This briefly describes the computations performed, regarding the determination of the Rewards rankings, and from 1998-2002, its top two teams agreed with the polls all ten times, though the order of the teams in 2002 was reversed – which would not change who would be playing in the BCS championship game. (A brief example regarding the 2024 Rewards final, pre-bowl rankings will also be provided now.)

Using the 2024 season, Table 1 holds the final Rewards ranking, with ratings, where: the NO_MOV column holds the NO_MOV rating for that team; the next two columns hold the weighted average win value, the sum of all losses, and the final two columns ensure that a team that plays an additional game or two is not rated higher than other teams. (In this case, all teams played at least twelve games, and so a win over the weakest teams played will be ignored if a team played more than twelve games.)

Table 1: Rewards Ratings for Its Top Eight Teams in 2024

Rank	Team	W	L	Rating	NO_MOV	Win Value	Loss Value	'W'	'L'	CFP
1	Oregon	13	0	10.0546	0.9963	0.84	0.00	12	0	1
2	Georgia	11	2	7.7572	0.8709	0.83	-0.51	10	2	2
3	Ohio State	10	2	7.5513	0.8819	0.78	-0.28	10	2	6
4	Notre Dame	11	2	7.0377	0,8893	0.69	-0.56	10	2	5
5	Indiana	11	1	6.9240	0.8168	0.64	-0.12	11	1	7
6	Boise State	12	1	6.7298	0.7789	0.61	-0.00	11	1	8
7	Texas	11	2	6.5518	0.8648	0.68	-0.26	10	2	3
8	Penn State	11	2	6.5488	0.8528	0.67	-0.12	10	2	4

The NO_MOV ratings for those twelve games, which were used to determine the average win value, and the loss value, for the top eight teams in 2024 appear, in Table 2. (All of these values store more digits of precision internally, in the software that works with them, however, those values display less precision, so that they could fit into Table 2's columns. So, to compute Oregon's Rewards rating, you would add 0.882 * 1 (a win over Ohio State) + 0.853 * 1/2 (a win over Penn State) + 0.779 * 1/4 (a win over Boise State, and so on) + 0.743 * 1/8 + … + 0.264 * 1/2048 = 1.67565 and then divide this by 1 + 2047/2048 (which almost equals 2) yielding roughly 0.84 (though the *correct value* would be used in the denominator, in the software). Finally, any losses would be subtracted after multiplying by 'W', which in this case yields 10.05. Georgia defeated Texas twice, and that is why 0.865 appears twice – as do the losses to Georgia for Texas (-0.13). Losses to Oregon, for Ohio (and Penn) State each subtracted 1- 0.9963 = 0.037 from these multiplicative results, which is why you see -0.00 in those rows below. Ohio State's actual average win value is less than 0.78, so when you multiply that true value by 10, and then subtract 0.28, the Reward rating is 7.5513. Likewise for Penn State, whose average is slightly smaller than the 0.67 displayed above; after multiplying that quantity by 10, even though Penn State won 11 games in 2024, and subtracting 0.12, their Rewards rating is 6.5488.

Table 2: Individual Game No_Mov Ratings for Top Eight Teams In 2024

Team	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th
Oreg.	0.882	0.853	0.779	0.743	0.725	0.586	0.560	0.500	0.486	0.441	0.367	0.264
Geog.	0.865	0.865	0.767	0.733	0.648	0.647	0.440	0.436	0.298	0.120	-0.29	-0.22
Ohio	0.853	0.817	0.653	0.649	0.533	0.500	0.423	0.366	0.293	0.264	-0.28	-0.00
Notre	0.720	0.680	0.680	0.648	0.600	0.582	0.490	0.488	0.362	0.328	0.264	-0.56
Indi.	0.725	0.586	0.560	0.533	0.500	0.441	0.423	0.341	0.264	0.096	0.063	-0.12
Boise	0.649	0.649	0.564	0.506	0.432	0.367	0.283	0.231	0.222	0.201	0.135	-0.00
Texas	0.725	0.680	0.647	0.582	0.546	0.490	0.482	0.436	0.338	0.327	-0.13	-0.13
Penn	0.743	0.606	0.600	0.586	0.560	0.529	0.486	0.441	0.440	0.264	-0.12	-0.00

In the paper by Mease he compared his approach with all of those former published BCS rankings – and the BCS computer systems included in each year's BCS ranking. For the final polls released between 1998 and 2002, that system (by Mease) had 8 correct (and 2 incorrect) teams matching who the poll's listed as the top two teams. The two computer system rankings produced by Sagarin, and Anderson and Hester, had 7 correct (and 3 incorrect), and all the other systems were not in use for all five years; however, all of them missed at least of one of the top two teams [4].

Three system's rankings were present in four of those years; the system by Massey was 7-1, and the other two were 6-2. Six other systems were included in the study by Mease and the three that were only present (in the BCS formula) for two years went 3-1; the three other systems had two incorrect matchings with the polls' top two teams. (Please remember that the top two teams in the polls are being compared in these results, which in some years, were different from the two teams who played in the BCS championship game, since the BCS formula occasionally did not agree with the poll's two teams) [5].

In subsequent years of the BCS, the Rewards system performed reasonably well, matching the top two teams five of the next eleven years, and only in 2008 did it miss both top two teams (in the polls) being ranked in its top two. Overall, it matched 25 of the top two teams, over those 16 years; consulting the PRS ranking, that system matched just 19 of the 32, as did the NO_MOV version – though those two systems agreed (with each other) on just nine of those 19 teams.

Transitioning to the College Football Playoff (CFP)

During the 16 years when the BCS was being used (especially after 2002), there were several years where it was quite clear that there were more than two teams who were worthy of consideration for being crowned that year's champion. Therefore, beginning in 2014, the CFP replaced the BCS and four teams would now be invited. The CFP committee would release its top 25 ranking each week, from about the middle of the season – until its conclusion. The results from many analytical tools were at the committee's disposal, to help guide their deliberations – when they applied the specific CFP rules set forth for arriving at said rankings – and one of the most important rules was one that is based upon the head-to-head competition/comparison of teams who played against each other that year (when present). This rule would be enforced when team A defeated team B and both teams' overall records were identical: team A would then be ranked higher/above team B. (This rule somewhat resembles the Rewards ranking system's rationale for weighting a team's victories exponentially – thereby allowing for significantly more weight to be given to wins over the strongest teams each team defeated that year, as determined by the NO_MOV ratings for all teams).

Increasing the number of invited teams from two to four was supposed to alleviate some of the controversy when only two teams were invited, which was inherent in the BCS selection process – especially when the decision regarding who the top two teams were wasn't obvious; 2014 provided plenty of controversy, especially with regards to the fourth and final team that was invited. No one seemed too concerned with the top three teams that were invited/selected; the controversy in 2014 was concerned with which of other three, once beaten teams would become the fourth team invited to that championship playoff/tournament.

Florida State (FSU) was the only undefeated team in 2014 – while Alabama, Oregon and Ohio State had earned records of 12-1; the conference Baylor and TCU belonged to did not have a season-ending, conference championship game that year, and so they were both 11-1. Since Baylor had defeated TCU earlier that year, they owned the head-to-head over TCU, but some people were confused when TCU was ranked 3rd, and above Baylor, in the penultimate CFP committee ranking – where Ohio State was 5th, one behind FSU and one above Baylor. This was mainly due to TCU having already defeated (then 9-2) Kansas State, who was to play Baylor in the final week of the season (when said conference championship games would be played as well). Some fans who did not realize that a win by Baylor would catapult them back in front of TCU were confused when that actually did occur, and, Baylor fans were also shocked when Ohio State also remained above them in that final ranking. (Ohio State did go on and win both of its playoff games, and was therefore declared as the National Champion in 2014; given that outcome, they were certainly worthy of that final invitation.) Table 3 illustrates how close these six teams were – via the values produced by the three computer systems described herein – and the Rewards system, because of its strategy to reward heavily any wins over strong teams, weighed the Baylor win over TCU significantly enough to have them ranked fourth in its final order (which is what all subscripts depict – in Table 3).

Table 3: Some Quantitative Performance Results for the Top Six Teams in 2014

2014	CFP	Polls	PRS	NO_MOV	Rewards
Alabama	1	1	133.07_1	101.14_2	8.81_3
Oregon	2	3	128.41_3	101.13_3	9.28_1
Florida St	3	2	117.001_6	101.19_1	9.26_2
Ohio St	4	5/4	126.36_8	101.02_4	8.010_6
Baylor	5	4/5	125.50_9	100.84_9	8.51_4
TCU	6	6	132.21_2	100.93_6	8.017_5

Other Possible Objective Models for Determining the Best Teams

In light of this (somewhat) controversial choice of Ohio State over Baylor, the possible strategy of rewarding teams for wins over strong teams differently (than how that is accomplished in the Rewards system) – as well as penalizing teams more for losses against weaker opponents – was investigated to determine if the committee's top four might actually have *had* the best seasons in 2014. Two ideas were combined to implement many models that would support a new approach. The first idea was to break the entire set of Division 1 teams into groups where each group's size decreased as the selection criterion went from the weakest to the strongest teams. Wins over teams in the biggest/weakest group earned zero points and then the points awarded for wins in each subsequent smaller group would increase since those groups should hold teams whose performance that year was *above* those in the larger group *below* it. Eight group selection policies were deemed reasonable (and relatively different from each other) and four, point value awarding policies were also devised with increases that were: linear (L), quadratic (Q), similar to the Fibonacci series (F), and exponential (E). This meant that there were 32 objective, model combinations in all, when excluding MOV; however, another 32 could be evaluated when including the full MOV as well (creating 64 objective evaluation models in all). Embedded Tables 4a and 4b list the four increasing, victory award, point value sequences as well as the eight different set sizes of *performance comparable* teams (according to the rules that allocate those groupings). These patterns mostly follow decreases according to the fractions listed in Table 4b, or one can also see the patterns with the decreasing percentages present here. Group G uses the *Golden Ratio* to reduce each set by roughly 61.8%, as derived by $1 / \phi$, where is $\phi = (1+\sqrt{5})/2$ and is roughly equal to 1.618, so $1/ \phi$ is roughly 0.618. (One can also algebraically prove that $\phi - 1/\phi = 1$.)

Tables 4a and 4b: Specific Parameter Values for When The 32 Objective Models are Evaluated

Table 4a: Point Values (PV); Group Sizes in Table 4b below vary from using 4 to 7 of these PVs							
Group Size	Largest	Points awarded when opponent resides in this particular group					Smallest
E	0	1	3	4	8	16	32
F	0	2	3	5	8	13	21
L	0	1	2	3	4	5	6
Q	0	1	3	6	10	15	21
Table 4b: Group Sizes (from Weakest to Strongest; the Fractions shown Illustrate the Pattern)							
E	0.5	0.25	0.25	0.125	0.0625	0.0625	-------
F	0.4	0.3	0.2	0.1	-------	-------	-------
G	0.375	0.2625	0.1625	0.10	0.6	0.4	-------
H	0.25	0.25	0.2	0.15	0.10	0.05	-------
N	1/4	1/5	1/6	1/7	1/8	~1/9	-------
O	1/3	1/5	1/7	1/9	1/11	1/13	~0.45
V	0.35	0.3	0.2	0.1	0.05	-------	-------
W	1/3	1/4	1/5	1/6	1/20	-------	-------

For the 32 objective models that ignored MOV – after establishing the respective, eight group boundaries, once all Division 1 teams had been ordered essentially by their W/L records, and by some measure of their SOS that year – 24 of these models agreed with who the CFP committee ranked as the top four teams (though not necessarily in exactly the same order; many of them reversed Oregon and FSU). More can be found about this study which also includes results from these 32 (or 64, overall) objective models with respect to the 2015 season as well [6].

The initial, high-level results of this study therefore helped to somewhat lessen the controversy, given the objective nature of the models that were described and evaluated; a second study extended this investigation through the 2018 season (five years of CFP results) – which is when a different approach was investigated. Table 5 illustrates the performance levels for both types of systems, and summarizes how many of the four top teams, in the five years from 2014-2018, had been correctly chosen by these 64 models (from those 20 invited teams, which is the maximum value that could appear in any cell inside Table 5's bottom two rows) [7].

Table 5: Number of Top Four CFP Teams as Selected by the Models from 2014-18

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

Table 5 highlights the effectiveness for all of the devised, objective models with regards to matching the CFP committee's selections. Over the first five years of the CFP, both WL models matched 18 of the 20 teams that were selected by the CFP committee. (The grouping policy precedes the point value sequence in the two letter model names used throughout this paper, with regard to these 32 objective models – both with and without MOV.) One can also observe, that on an overall basis, the models including MOV have slightly outperformed those that exclude MOV, averaging roughly 15.2 correctly matched teams versus approximately 13.9 teams. (WF and VL were the only two models that matched 17 correctly – both with and without MOV. So, both of the most accurate models listed here reward wins over slightly more than 30% of the teams – which reside within the weaker segment of the performance spectrum that year – zero points towards a team's final total.)

To avoid any possible bias being introduced, via the initial ordering of all of those teams into groups, one million random permutations of all Division 1 teams were used to arrive at each model's final ranking where one permutation's produced ranking would then be used to set the next iteration's initial group assignments, and then another ranking would be generated from those assignments and so on, until the ordering produced had converged to the ranking of the previous iteration (from that initial, random permutation). (Typically, nine to fifteen iterations are required before convergence occurs. These final, convergent rankings were used to generate Table 5 above.)

This approach was somewhat personally unsatisfactory mainly because some of these objective models *will* perform better than others, and one can then simply *use* the most accurate model. The success of these models was somewhat unfulfilling because any set of strongly matching results relies solely upon the ideas that established the particular point value and group selection choices. However, a completely new approach became evident after presenting these models that precipitated another possible way to decide which four teams the CFP committee might select.

The Improved Linear Model (ILM)

During the conference where the article describing the successes of the approach outlined in the previous section was presented, the idea of examining how well the basic PRS and NO_MOV systems matched the CFP committee's choices from 2014-2019 came to the forefront. As it turns out, the PRS correctly identified 16 of the 24 teams chosen in the CFP's first six years, and NO_MOV did so for 20 of the 24 teams, where 15 of those invited teams were in both systems' top four teams. So, the idea was that after computing those two quantities, perhaps a linear combination of those two ratings could be found that would match the committee's ranking more closely than either one separately, instead of trying to determine which of the aforementioned, objective computer models mimics most closely that committee's final ranking [7]. To be succinct here, since this is explained in greater detail elsewhere in both of those power rating systems, PRS and NO_MOV, a team's rating is the sum of its strength of schedule (SOS) and the average difference between the offensive and defensive point totals for each game (for that team) [8]. The formula to determine a team's CFP value, in this *Improved Linear Model* (ILM), would need a set of five weights – to be multiplied individually against the four values mentioned above; the 5th weight would be multiplied by a team's losses and that value would be subtracted (in a similar manner to the original BCS formula).

After employing the Monte Carlo simulation technique to generate one million random sets of five weights, several dozen, different sets performed at the same level where nine teams were ranked in exactly the same position by the ILM as was done by the committee, and overall, fourteen of the sixteen invited teams in the training data set (2014-2017) were also in the ILM's top four. (This performance will be denoted as (9,14) for the remainder of this article.) To settle on one set of weights, the set that maximized the Spearman Correlation Coefficient (SCC) for the CFP's top 25 teams, for all four years in the training set, was to be used to make predictions for the years that followed those first four years (of the CFP). In the next two years, using those five weights, the ILM matched the top eight CFP committee teams exactly, in 2018, and in 2019, it matched all four top teams – though the top two teams were in the opposite order as chosen by the committee. (The weights for the two PRS rating components were uniformly chosen to be between zero and one, uniformly between zero and ten for the loss weight, and uniformly between zero and 100 for the NO_MOV rating components; see [6] for more details as to why those ranges were chosen). Therefore, the ILM matched 22 of the 24 invited teams in the CFP's first six years; that performance is better than the individual combined ratings of PRS (16) and NO_MOV (20). (Excluding the first four years of training, ILM matched all eight invited teams, whereas PRS and NO_MOV both only matched three out of four, in 2018 and 2019, for a total of six out of eight.)

Because the COVID pandemic forced many colleges to restrict its teams from travelling and/or competing against other schools outside of their own conference, the 2020 season will not be considered over the rest of this paper. In 2021, the ILM was (2,4); in 2022, it was also (2,4), with the #2 and #3 teams in the opposite order – which would not impact which teams would be competing in the first round of the CFP – and in 2023, it had three of the top four teams chosen by the CFP but none in the same rank as the committee had deemed appropriate, i.e. (0,3). So, for the five years (2018-2019, 2021-2023) that came after the four years that were used for ILM training purposes (2014-2017), this model only missed one invited team, and it also match the committee's position nine times, i.e. (9, 19), for the twenty teams the committee invited in those five years. Therefore, over all nine years, the ILM was (18, 33), missing only three of the 36 actually invited teams (by the CFP committee), whereas the Rewards system was (11, 28); excluding the first four training data years, Rewards yielded (7, 15).

WL and NL were the top two models – both using PRS or NO_MOV – for matching the committee's choices in 2014 and 2015. For 2014-23, WL – both using and excluding MOV – is still the best at matching 32 of the 36 top four teams chosen over those nine years, where both WF models were second best (31), and several other models achieved 30 out of 36. (The ILM also matched 26 of the 32 BCS teams selected from 1998-2013 as well as matching 26 of the top two teams in the polls during that same sixteen-year span. The yearly ILM rankings can be found online if you expand the Improved Linear Model line beneath the Current Research Projects section heading at https://www.smcvt.edu/directories/employee-directory/john-trono).

The Expansion of the CFP

In 2024, to avoid the controversy of only having four teams compete for the right to be crowned as the National Champion, the invited field was expanded to a dozen teams, and the top twelve teams in the final CFP committee ranking would be invited *unless* the five, highest CFP ranked conference champions were not already present in the top twelve. (In 2024, only one such team resided outside the committee's top twelve, and so the #11 team, Alabama, was replaced by Clemson, who was the ACC conference champion; the #12 team was Arizona State (ASU) – which remained in the CFP because it was representing the Big 12 conference since ASU had won that conference's championship game.) So, the question then arose: how well would the ILM perform in 2024 given that it was trained to match the committee's top four – not top twelve – positions in that final ranking? And, even if it performed reasonably well, could there also exist another set of weights that would make more accurate committee predictions of the committee's top twelve?

The original ILM did reasonably well when retrospectively applied to those first nine years of CFP invite matching 31 teams – and their positions – exactly, and 96 of the overall 108 teams in the top twelve: i.e. (31,96). (This assumes that the committee's previous rankings would be the same if they had been previously tasked with inviting twelve – and not just four – teams).

Employing Monte Carlo, simulation once again, fifteen billion random sets of five weights were generated, using the same, designated ranges as previously described, and many of those results can be found in the conference paper or in the expanded online version; the latter contains an extensive, five-page Appendix that had to be omitted from the conference paper, which could be at most six pages [9,10]. To reiterate some of those outcomes succinctly here, there were four new sets of weights that were chosen for their performance, with ties amongst models with the same performance outcomes once again broken by choosing the one with the highest SCC value for the top 25 teams (SCC-25) that appeared in the committee's final ranking during the nine years in the training data set. These models have simply been named by their performances – (33,100), (38,99), (40,98) and (41,96) – by excluding the parenthesis, and several other such models were derived as well. A set of weights that yielded the highest SCC-25 over the nine years of training data (2014-19, 2021-23), for this expanded investigation, is also included in an upcoming table (under the name **SCC**).

Also, a previous study only investigated matching the top four teams that were deemed worthy of being invited (by the CFP committee), so two new formulas were devised to provide two more quantitative measures to evaluate a weight set's performance – besides the number of exact and overall matches in the top twelve CFP selections [8]. These formulas assign a rewarded point value for each rank; the closer the prediction is to the actual CFP rank, the larger the portion of this point value that will be earned. In these formulas, it also seemed more important to match the higher ranked teams, and so exactly matching the CFP #1 team would be worth 100 points, with regards to these two performance measures, whereas exactly matching the #12 would only be worth 34 points.

The integer quantities in the top row of Table 6 are used to depict both the CFP ranking position for a team, as far as how many points that position would be worth when matched exactly, as shown in the second row of this table, as well as representing the difference between a team's CPF ranking and where they were placed by an ILM model with regards to the values that appear in the third and fourth rows in Table 6 – with regards to the following two (named) formulas: F_Sqr and F_*. The values in the second row of Table 6 decrease in a linear fashion, starting with subtracting eleven and then ten – all the way down to a one-point difference between the full reward for matching the teams ranked eleventh and twelfth (by the CFP committee). The first formula mentioned above (F_Sqr) is: max(0, VCR – (actual team's CFP rank – the model's rank)^2), where VCR stands for the value for that team's CFP ranking. Likewise, the second formula (F_*) is: max(0, VCR – VCR for that CFP ranking * (actual team's CFP rank – the model's rank)/10.0)). The quantities in the 3rd and 4th rows, in Table 6, represent how much the second row's point value award would be reduced by, when applying the two formulas for that specific difference that appears in the top row, in that column.

Therefore, according to the two formulas in the preceding paragraph, the value earned per rank would be reduced by the square of the difference between the weight set's prediction and the actual CFP rank in the first formula (F_Sqr). The other formula, F_*, will reduce the value earned as well – but in this case, the reduction is multiplicative in nature; for each position that the prediction is further away from the actual CFP rank, the point value reduction is by another 10%. So, for example, if the prediction for the team ranked as the fourth team is off by three, the value awarded for the prediction produces an award of 61 points (70 – 9) when using the F_Sqr formula, whereas F_* would reduce the award by 30% yielding 49 points (70 – 70 * 0.3). The bottom two rows in Table 6 illustrate the reduction that would occur for the differences appearing in the top row in Table 6.

Table 6: Point and Update Values for these Two Formulas

	1	2	3	4	5	6	7	8	9	10	11	12
Value	100	89	79	70	62	55	49	44	40	37	35	34
F_Sqr	-1	-4	-9	-16	-25	-36	-49	-64	-81	-100	-121	-144
F_*	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.0	1.0

These two measures can also be used as a criterion for choosing a weight set and so F_Sqr and F_* will be used both as the name for an evaluation measurement as well as the name of the weight set that maximized that quantity for all generated weight sets during the Monte Carlo generation (with the training data set). Both of these formulas will award a minimum of zero points if the difference between the prediction and the actual rank is large enough; for F_*, this difference would be 10 or more places – and for F_Sqr, the largest difference that still produces a positive point value depends upon the CFP rank of the team in question. The results, with regards to the training data set (2014-2019, 2021-2024), were (32,99) for F_*, and (30,99) for F_Sqr, and, SCC achieved (28,96) over those nine years.

Of the four newly named models, 38 + 99 > 33 + 100, and even though 40 + 98 > 41 + 96 = 38 +99, there was a worry that the two sets of weights that matched more than 39 team's positions exactly could be overfitting the results, and so it was conjectured that the weights for the model *named* 38,99 might be the *best* performer in 2024. (All of the weights for all of these models were determined before the end of November, 2024, and the final CFP committee ranking was released in early December).

How well all eight models performed (in 2024) appears in Table 7, where the model names are listed in the top row, and the types of performance measurements are listed in the leftmost column in this table. It was a little surprising that the original ILM weight set achieved the highest SCC-25 value, and it seems that the weights associated with the 38,99 model had the best results since it was the only one to have eleven teams in *its* top twelve along with three exact matches of the CFP committee's final ranking, i.e. 38, 99 achieved (3,11). The model named 40, 98 also performed well, achieving (2,11). The model named 38,99 also had an only slightly smaller SCC-25 value (than the original ILM weighted model) as well as it also being associated with the largest F_Sqr, F_* and SCC-12 values (of the eight models), which is not so surprising given its strong performance in 2024.

Table 7: Actual Results for the 2024 NCAA Season

2024	Orig.	33,100	38,99	40,98	41,96	SCC	F_Sqr	F_*
Results	1,11	1,10	3,11	2,11	1,10	2,10	2,10	2,10
SCC-25	.8796	.8231	.8738	.8642	.8565	.8262	.8273	.8304
F_Sqr	587	588	619	586	582	576	591	591
F_*	555.0	525.1	553.9	530.7	524.6	559.0	563.8	563.8
SCC-12	.6259	.6294	.7663	.6538	.6399	.5874	.6399	.6399
SCC-4	-0.1	-0.1	-1.8	-1.6	-1.4	0.0	0.0	0.0

When comparing these model's rankings, a curious observation was that for all eight of them, the CFP committee's #4 team (Penn State) and #10 team (SMU) always appeared as the 7th and 8th rated team for each model. (Oregon was the CFP committee's #1 team and they were also ranked #1 by all eight models as well.) Table 8 lists the integer, positional ranks for all of these models whose weights have been recently created (along with the original ILM weights). One might expect that the F_Sqr results to be roughly the same as the results that maximized the F_* formula, and in Table 8, only Missouri and Colorado are not identical.

Table 8: Integer Model Ranking Values for all Teams Ranked 25th – or Higher

CFP	Name	Orig.	33,100	38,99	40,98	41,96	SCC	F_Sqr	F_*
1	Oregon	1	1	1	1	1	1	1	1
2	Georgia	3	3	5	6	6	2	2	2
3	Texas	4	4	3	4	4	4	4	4
4	Penn State	7	7	7	7	7	7	7	7
5	Notre Dame	2	2	2	2	2	3	3	3
6	Ohio State	5	5	4	3	3	5	5	5
7	Tennessee	15	13	11	11	11	14	13	13
8	Indiana	6	6	6	5	5	6	6	6
9	Boise State	11	15	14	14	14	15	15	15
10	SMU	8	8	8	8	8	8	8	8
11	Alabama	12	9	9	10	9	10	9	9
12	Arizona State	9	11	12	12	13	9	10	10
13	Miami (F.)	10	12	10	9	10	12	12	12
14	Mississippi	18	17	16	16	15	18	17	17
15	South Carolina	13	10	13	13	12	11	11	11
16	Clemson	16	16	17	17	17	16	16	16
17	BYU	14	14	15	15	16	13	14	14
18	Iowa State	17	18	18	18	18	17	18	18
19	Missouri	21	20	23	23	23	20	21	20
20	Illinois	20	24	24	24	24	22	24	24
21	Syracuse	19	23	25	25	25	21	23	23

22	Army	23	26	21	20	22	25	25	25
23	Colorado	24	21	19	19	19	24	20	21
24	UNLV	30	34	30	30	32	34	34	34
25	Memphis	36	38	36	36	36	38	38	38
NR	LSU	22	19	20	22	21	19	19	19
NR	Louisville	25	22	22	26	26	26	22	22
NR	Texas A&M	26	25	26	28	29	27	26	26

One More Set of Models (Devised from Observations After the 2024 CFP Ranking)

Since the ILM is a linear combination of the PRS and NO_MOV ratings, that are computed from all the games played in a particular year, it seems to have been impossible for Tennessee and Boise State to both be in the top twelve teams for any set of five weights to determine their ILM rating/ranking (after the 2024 season was concluded, and before any bowl games/CFP matchups were played) in the linear manner just described. In hindsight, this seems obvious since Tennessee was ninth in the PRS ranking, and just nineteenth in the NO_MOV system where Boise State was ranked tenth there, and 33rd in the PRS ranking. (Another Monte Carlo simulation was performed, generating weights in an attempt to see if these two, top twelve CFP ranked teams could both be placed there, via this approach; no such weights were found in the one billion sets that were examined).

Revisiting the earlier comment regarding the CFP #4 team (Penn State) always being ranked as the #7 team in all of these *linear*, ILM-style models, one final thought experiment seemed worth investigating. After modifying the Rewards system by eliminating the weights for the wins, thereby counting each win *identically*, then, instead of rewarding each win with the opponent's NO_MOV rating, could there exist a simple, arithmetic combination of the normalized (between zero and one) PRS and NO_MOV ratings that might match more closely those top twelve, committee chosen teams? When just adding PRS and NO_MOV as the reward for defeating a team, the top three CFP teams appeared in those same positions, though Penn State only moved up to #5.

However, when multiplying those two ratings, to be the reward for a win over a team, Table 9 illustrates what this approach generated. Therefore, even though the ILM approach seemed to always place Penn State as its seventh highest ranked team, this updated, Rewards-style strategy did match the top four teams, in the committee's final ranking, and, both Tennessee and Boise State were also in the top twelve for this new approach for rewarding wins (and losses). (This model is listed as 'MU0' in Table 11; the other acronyms/approaches listed there will be described briefly before analyzing the results when evaluating these combinations. MU0 was the only one of these twelve models, in Table 11, to correctly predict all four of the 2024, top four committee choices – as well as having all four in the same order as the CFP committee.)

Table 9: Output Produced by the Modified Rewards System Program (MU0)

Rank	Team	CFP	W	L	Rating
1	Oregon	1	13	0	4.7784
2	Georgia	2	11	2	4.2490
3	Texas	3	11	2	3.1703
4	Penn State	4	11	2	2.8348
5	Ohio State	6	10	2	2.7705
6	Notre Dame	5	11	1	2.7700
7	Indiana	8	11	1	2.2790
8	SMU	10	11	2	2.2767
9	South Carolina	15	9	3	2.1556
10	Alabama	11	9	3	2.0688
11	Tennessee	7	10	2	1.9916
12	Boise State	9	12	1	1.9112
13	Clemson	16	10	3	1.8114
14	Miami (F.)	13	10	2	1.7636
15	Arizona State	12	11	2	1.7159

(One might recall that Ohio State and Notre Dame were ranked as the #3 and #4 teams in the Rewards ranking, as seen in Table 1, and since those two teams ended up playing for the 2024 National Championship, perhaps said rating strategy was a more accurate evaluation of all of these teams performances in 2024 than either the MU0 approach above, or those people sitting on this year's CFP committee.) To determine each team's MU0 rating, one must simply sum the twelve values that appear in Table 10.

Table 10: Teams in 2024; (Remember that the values below are the Opponent's NO_MOV Rating Times its PRS Rating)

Team	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th
Oreg.	0.861	0.720	0.543	0.513	0.485	0.355	0.326	0.288	0.251	0.229	0.134	0.073
Geog.	0.867	0.867	0.682	0.618	0.505	0.452	0.322	0.313	0.168	0.025	-0.31	-0.26
Texas	0.540	0.513	0.505	0.438	0.399	0.353	0.313	0.195	0.168	0.142	-0.20	-0.20
Penn	0.485	0.454	0.437	0.355	0.326	0.288	0.278	0.229	0.204	0.073	-0.16	-0.14
Ohio	0.764	0.720	0.474	0.383	0.345	0.251	0.183	0.129	0.091	0.073	-0.49	-0.16
Notre	0.611	0.540	0.455	0.454	0.452	0.034	0.262	0.168	0.155	0.073	0.061	-0.80
Indiana	0.513	0.355	0.345	0.326	0.251	0.229	0.183	0.111	0.073	0.027	0.003	-0.14

Because both normalized ratings are less than one, their product could be significantly smaller, and so it was deemed worth considering using the rating values as well as the value of the rating plus one in the formulas that would produce the amount each team would accrue (relying on one, or both, of the opponent's two ratings) for each of their wins (over each of their opponents). All possible combinations for each rating, both as is, or, with both ratings being incremented – via addition or multiplication operations (that seemed reasonable), with regards to both the PRS and NO_MOV rating systems – are listed in Table 11. The first two rows simply list models using just the NO_MOV or PRS rating, or incrementing them, before filling in the values as they appear in Table 10. The suffix of zero, at the end of the models in Table 11, indicate that just the ratings are used, where the models ending in '1' will have the value of 1 added to any rating that is used before addition or multiplication occurs. AD is for just adding both ratings, MU for multiplying them, AM adds the sum and product of both ratings whereas MM multiplies that sum and product (to yield that model's rating – for MM0 and MM1).

Table 11: List of Twelve PRS and NO_MOV Combined Reward Strategies (For Each Win)

P10 = (NO_MOV)	P11 = (NO_MOV + 1)
PW0 = (PRS)	PW1 = (PRS + 1)
AD0 = (NO_MOV + PRS)	AD1 = (NO_MOV + 1 + PRS + 1)
MU0 = (NO_MOV * PRS)	MU1 = ((NO_MOV + 1) * (PRS + 1)
AM0 = (NO_MOV + PRS) + (NO_MOV * PRS)	AM1 = ((NO_MOV+1) + (PRS + 1)) + ((NO_MOV + 1) * (PRS + 1))
MM0 = (NO_MOV + PRS) * (NO_MOV * PRS)	MM1 = ((NO_MOV + 1) + (PRS + 1)) * (NO_MOV + 1) * (PRS + 1))

Some of the formula above could have been reduced, but were not – so that the form of how all of them were created was apparent. (For example, the '+ 1' appearing twice in AD1, for instance, would appear as '+ 2' in the software for AD1. One would think that AD1 and AD0 would produce identical orderings, however, AD1 ratings are one higher than the AD0 rating for each win earned, which bolsters the final rating for those teams that had better W/L records – over those teams with more losses.) If one were to utilize the same nine years of data that the ILM models (for the expanded, twelve team, CFP) were *trained* on, the previous 64 aforementioned, objective models, and the twelve strategies in Table 11, produced the following results.

As one can see in Table 12, where the number of exact matches appears in the second row, and the total number of matches (of the CFP committee's top twelve, invited teams) appears in the third (and bottom row), there were a couple of these dozen models (from Table 11) that performed pretty well on the training data set. AD0 and MM1 achieved the highest sum (37 + 93 = 40 + 90 =130), though AD0 had the highest SCC-12 value (0.7104; MM1 was third) and the highest SCC-25 value (0.8443l; MM1 was fourth). In 2024, AD0 and MM1 had five of the top twelve in the exact same position as the committee, and both matched all but one of the committee's top twelve teams. (Nine of these dozen models matched eleven – and the other three matched ten – and only one, AM0, had six of the twelve ordered exactly as the committee did, and eleven of the twelve).

Table 12: New Model's Performance on the Training Data Set

P10	P11	PW0	PW1	AD0	AD1	MU0	MU1	AM0	AM1	MM0	MM1
22	25	34	29	37	30	27	36	34	32	29	40
93	85	93	88	93	87	90	89	90	89	91	90

With regards to the 64 objective models described earlier, WL still performed the best with regards to the committee's top four choices, matching 31, or 32, of the 36 invited teams over the CFP's first nine year (excluding 2020): (13,31) with MOV, and (13,32) without. HL was (15,32) with MOV and (17,30) without; HF was (18,29) with MOV, and NL was (19,29) without MOV. Regarding the 108 teams in the top twelve over those same nine years, WL did well – (21,93) with MOV, and (25,90) without – but some others were superior: HL was (30,96), using MOV, while HF was (28,94); without MOV, NL was (19,93). In 2024, for the models using MOV, fourteen of those 32 models matched eleven invited teams, and the rest matched nine or ten, (the other two only matched seven). Without MOV, none matched eleven, eighteen matched ten, twelve matched nine and two matched eight.

Summary

Many different objective strategies have been described here as well as how accurately they have performed - with regards to the manner in which teams have been invited to compete for the NCAA National Championship (in football). It seems unlikely that the human element will ever be completely removed from this selection process though it is satisfying to realize that there are (typically) several different, objective (and unbiased) methodologies that can confirm the validity of such selections as well as to help all of those involved feel justified that there are many quantitative measurements to verify that the invited teams did deserve that opportunity to complete for such a title.

Postscript

Table 13 below showcases how well each of the ILM models performed in 2025 (as well as the 2024 results, for comparison purposes) since the final, 2025 CFP rankings were released one week before this article was submitted.

Table 13: Actual Results for the 2024 and 2025 NCAA Seasons

Results	Orig.	33,100	38,99	40,98	41,96	SCC	F_Sqr	F_*
2024	1,11	1,10	3,11	2,11	1,10	2,10	2,10	2,10
2025	3,11	2,11	3,11	4,11	4,11	2,12	2,11	2,11

Table 14 does likewise for the newest models, that appear in Table 11. In 2025, P11 matched the 1st seven teams exactly (as well as #11); sadly, as illustrated in Table 12, the P11 model had the lowest performance results of said dozen models during the nine years in the training data set.

Table 14: Actual Results for the Twelve Models Illustrated in Table 11

Results	P10	P11	PW0	PW1	AD0	AD1	MU0	MU1	AM0	AM1	MM0	MM1
2024	5,10	1,10	1,11	2,11	5,11	1,10	4,11	4,11	6,11	2,11	4,11	5,11
2025	4,12	8,11	4,12	4,11	4,12	4,11	3,12	3,12	3,12	3,11	3,12	3,12

As you can readily observe in Table 15, even though P11 matched eight positions in the top twelve exactly, it's SCC value for the top 25 is also the lowest. The values in Table 15 that appear in bold are the best/highest for all twenty models contained therein. (P10 and P11, which both rely solely on the NO_MOV rating, matched the top four teams exactly, yet also had the lowest F_Sqr values.) Eight of the twelve models matched all dozen top teams, as placed there by the 2025 CFP committee, albeit in slightly different orders (than them); the other four models matched eleven of these top twelve teams.

Table 15: Collection of Measurements for all Twenty Models Appearing in Table 13 and 14

2025	Results	SCC-25	SCC-12	SCC-4	F_Sqr	F_*
Original	3,11	0.8904	0.8217	0.8	643	609.4
38,99	3,11	0.8746	0.7622	0.6	626	592.2
40,98	4,11	0.8738	0.7622	0.6	626	601.4
41,96	4,11	0.8638	0.7622	0.6	626	601.4
33,100	2,11	0.8996	0.8357	0.6	647	607.1
SCC	2,12	0.9073	0.8322	0.5	646	599.3
F_Sqr	2,11	0.8954	0.7832	0.5	632	591.2
F_*	2,11	0.8992	0.8322	0.5	646	599.3
P10	4,12	0.8635	0.7972	**1.0**	577	630.3
P11	**8,11**	0.7473	0.6923	**1.0**	595	**632.2**
PW0	4,12	**0.9177**	0.7622	0.1	628	603.4
PW1	4,11	0.8469	0.6993	0.6	615	598.7
AD0	4,12	0.8988	0.7413	0.1	622	595.7
AD1	4,11	0.7954	0.6818	0.6	612	595.0
MU0	3,12	0.8915	0.7413	0.1	635	603.8
MU1	3,12	0.8992	**0.8951**	0.6	**664**	629.3
AM0	3,12	0.8988	0.7762	0.1	632	602.3
AM1	3,11	0.8588	0.7727	0.6	629	609.3
MM0	3,12	0.8746	0.7752	0.1	632	602.3
MM1	3,12	0.8985	0.7972	0.6	646	586.0

Appendix

To illustrate how the power ratings (in the PRS) are computed, let's consider a fictitious set of six teams, where Team1 is ten points better than Team2, who is ten points better than Team3, and so on down to Team5 being ten points better than Team6. Assuming that these team strengths follow the rules of transitivity, then Team1 would be 50 points better than Team 6. And, there would be no loss in generality (in this case) if we also just assumed that Team1 always score 50 points, and Team2 always scored 40 points, all the way down to Team5 always scoring ten points, and Team6 never scoring at all (in any game).

Table A1: Average Margin of Victory for Those Six Fictitious Teams

W	L	Off	Def	OD	Games	Team
5	0	250	100	30.00	5	Team1
4	1	200	110	18.00	5	Team2
3	2	150	120	6.00	5	Team3
2	3	100	130	-6.00	5	Team4
1	4	50	140	-18.00	5	Team5
0	5	0	150	-30.00	5	Team6

Table A1 contains the won-loss records for each team, where every team played every other team once, along with the total points each scored (Off) and how many points their defense surrendered (Def). The OD column represents the average margin of victory which is simply the arithmetic mean of the points that team scored minus the arithmetic mean of the points that team surrendered. One of the main advantages of the power rating system is that it can deterministically calculate a relative strength of schedule (SOS) measurement that *makes sense* (as will be seen shortly). If you subtracted Team6's OD value from Team1's, this would lead us to believe that Team1 is 60 points better than Team6. However, Team6 played a more challenging, overall schedule because they played the best team, whereas Team1 played the worst team (given the four other opponents were the same for both of these two teams). Therefore, this OD metric is not sufficient to truly represent a team's ability; the power rating is simply the OD value plus the team's calculated strength of schedule, and here is how that SOS quantity is derived.

As just illustrated with Team1 and Team6, the expected point differential (between two teams that played against each other) is determined, and then compared with the actual score for all games (and for all teams) that season. Team1 would need to have -10 added to its OD quantity to produce the correct result (a 50 point win over Team6), and likewise -8, -6, -4 and -2 for the other four games Team1 played. That sum (-30) is divided by 5 producing a SOS value of -6.0 for Team1, that when added to Team1's OD value yields Teams1's power rating (24.0). The results for all six teams appear in the columns SOS1 and Power1 in Table 2. However, the same calculations (for all six teams) must be repeated, using the power rating instead of the pure OD value, to produce the next, estimated SOS value (and power rating). Table A2 shows that this takes five complete iterations before the power ratings for all teams have converged (and will not change anymore). One should also notice that the actual game scores are now more accurately (and exactly!) reflected in the power rating difference between competing teams [11].

Table A2: How the SOS and Power Rating Change (Until Convergence Occurs)

OD	SOS1	Power1	SOS2	Power2	SOS3	Power3	SOS4	Power4	SOS5	Power5	TM
30.00	–6.00	24.00	-4.80	25.20	-5.04	24.96	-4.99	25.01	-5.00	25.00	#1
18.00	-3.60	14.40	-2.88	15.12	-3.02	14.98	-3.00	15.00	-3.00	15.00	#2
6.00	-1.20	4.80	–0.96	5.04	-1.01	4.99	-1.00	5.00	-1.00	5.00	#3
-6.00	1.20	-4.80	0.96	-5.04	-1.01	-4.99	1.00	-5.00	1.00	-5.00	#4
-18.00	3.60	-14.40	2.88	-15.12	3.02	-14.98	3.00	-15.00	3.00	-15.00	#5
-30.00	6.00	-24.00	4.80	-25.20	5.04	-24.96	4.99	-25.01	5.00	-25.00	#6

References

1. Carlin B, Stern H (1999) Designing a College Football Playoff System. Chance 12: 21-26.
2. Berry S (2003) College Football Ratings: The BCS and the CLT. Chance 16: 46-49.
3. Trono J (2010) Rating/Ranking Systems, Post-Season Bowl Games, and 'The Spread'. Journal of Quantitative Analysis in Sports Volume 6: 3-6.
4. Mease D (2003) A Penalized Maximum Likelihood Approach for the Ranking of College Football Teams Independent of Victory Margins. The American Statistician 57: 241-248.
5. Trono J (2007) An Effective Nonlinear Rewards-Based Ranking System, Journal of Quantitative Analysis in Sports 3: 2, 3.
6. Trono J (2016) Is it Possible to Objectively Generate the Rankings Produced by the College Football Playoff Committee. Saint Michael's College https://www.smcvt.edu/wpcontent/uploads/2025/09/CFP_And_Partition.pdf.
7. Trono J (2019) Objectively Modelling the College Football Playoff Committee's Selections. Proceedings of the 7th International Conference on Mathematics in Sport https://www.smcvt.edu/wp-content/uploads/2021/08//CFP_Five_Years.pdf.
8. Trono J (2020) An Accurate Linear Model for Predicting the College Football Playoff Committee's Selections. Technical Report https://www.smcvt.edu/wp-content/uploads/2021/08/AccurateLinearCFPModel.pdf.
9. Trono J (2025a) Evaluating the Improved Linear Model (and its successor?) with regards to the expanded College Football Playoff. Proceedings of the 11th International Conference on Mathematics in Sports https://mathsportinternational.com/MathSport2025Proceedings.pdf.
10. Trono J (2025b) Can the Improved Linear Model be improved upon for the expanded College Football Playoff?. Technical Report https://www.smcvt.edu/wp-content/uploads/2025/05/ImprovedLinearModel_ExpandedCFP.pdf.
11. Carroll B, Palmer P, Thorn J (1988) The Hidden Game of Football, Warner Books.
12. Colley W (2002) Colley's Bias Free College Football Ranking Method: The Colley Matrix Explained. https://www.colleyratings.com/matrate.pdf.
13. Massey K (1997) Statistical Models Applied to the Rating of Sports Team. https://www.masseyratings.com/theory/massey97.pdf.

14. Stern H (1995) Who's Number 1 in College Football?... And How Might We Decide?. Chance 8: 7-14.