

An Accurate Linear Model for Predicting the College Football Playoff Committee's Selections

John A. Trono
Saint Michael's College
Colchester, VT
Tech report: SMC-2020-CS-001

Abstract. A prediction model is described that utilizes quantities from the power rating system to generate an ordering for all teams hoping to qualify for the College Football Playoff (CFP). This linear model was trained, using the first four years of the CFP committee's final, top four team selections (2014-2017). Using the weights that were determined when evaluating that training set, this linear model then matched the committee's top four teams exactly in 2018, and almost did likewise in 2019, only reversing the ranks of the top two teams. When evaluating how well this linear model matched the committee's selections over the past six years, prior to the release of the committee's final ranking, this model correctly predicted 104 of the 124 teams that the committee placed into its top four over those 31 weekly rankings. (There were six such weeks in 2014, and five in each year thereafter.) The rankings of many other, computer-based systems are also evaluated here (against the committee's final, top four teams from 2014-2019).

1. Introduction

Before 1998, the National Collegiate Athletic Association (NCAA) national champion in football was nominally determined by the people who cast votes in the major polls: the sportswriters' poll (AP/Associated Press), and the coaches' poll (originally UPI/United Press International, more recently administered by the *USA Today*). With teams from the major conference champions being committed to certain postseason bowl games prior to 1998, it wasn't always possible for the best teams to be matched up against each other to provide further evidence for these voters. The 1980s had three years ('82, '86 and '87 – as well as '71, '78, '92 and '95) where the top two teams in the polls competed against each other in a major bowl game – thereby *crowning* the national champion; however, many times, who deserved to be recognized as the national champion was not as clear as it should be.

For instance, there were three undefeated teams, and three more with no losses – and only one tie – marring their seasons, in 1973, and, in 1977, there were six teams (from major conferences) with only one loss after the bowl games were played. Who rightly deserved to be national champion at the end of those seasons? The two aforementioned polls reached different conclusions after the 1978 season ended: the AP pollsters voted Alabama #1, after they beat the then #1, undefeated Penn State team – while the coaches chose Southern California, who defeated Alabama 24-14 earlier that year (at a neutral site), but who later lost on the road to a 9-3 Arizona State team.

No highly ranked team – who wasn't already committed to another bowl game – remained to play undefeated BYU, after the 1984 season concluded, leaving 6-5 Michigan to go against the #1 team that year. In 1990, one team finished at 10-1-1 and another 10-0-1, and those two were obligated to play in different bowl games just like the only two undefeated teams left in 1991 – who could not meet on the field to decide who was best, due to their conference's commitments to different bowl games.

Perhaps the controversy that occurred in 1990 and 1991 helped motivate the NCAA to investigate creating a methodology to rectify this situation around 1992 (#1 did play #2 that year), eventually resulting in the implementation of the Bowl Championship Series (BCS), that began in 1998. Even though the BCS approach did select two very deserving teams to compete for the national champion each and every year that it was in place, during roughly half of those 16 years, it actually wasn't always clear if the two *best* teams had been selected – especially when there were between three and six teams some years whose performance during these particular seasons had provided enough evidence that those teams could've also been representative candidates to play in said championship game.

The College Football Playoff (CFP) began in 2014 (concluding the BCS era). To eliminate some of the controversy in response to the particular BCS methodology in use at that time, a reasonably large CFP committee was formed, whose constituency does change somewhat each year. The task this committee has been assigned is to decide who are the four *best* teams in the Football Bowl Subdivision (FBS) of college football that year (the FBS was previously called Division 1-A); the committee's #1 team will play the #4 team in one semifinal contest while the teams ranked #2 and #3 will play each other in the other semifinal, with the winners then meeting in the CFP national championship game.

2. Background

It is not difficult to find online many different approaches that *determine* which four NCAA football teams were the *best* that season. Rating systems will calculate a value for each team, and these systems are typically used to predict how many points a team will win by, against another team – on a neutral site. (The teams with the four highest ratings would then be *the best*.) These approaches utilize some function of the actual margin of victory (MOV) for each contest (if not incorporating the entire, actual MOV). Ranking systems tend to ignore MOV – only relying on the game outcomes – to order all the teams from best to worst.

If one were to rely on the ESPN Football Power Index (FPI) rating system to predict the CFP committee's choices, 15 of the 24 teams chosen, between 2014 and 2019, would've been correctly selected. (Notable omissions were the #3 seeded, 13-0, Florida State team, in 2014, who was ranked #10 by the FPI, and the #3 seeded, 12-1, Michigan State team, in 2015, who FPI ranked as #14.) Unlike almost all rating/ranking strategies, that rely solely on the scores of every game that was played that season, the Massey-Peabody Analytics group have utilized a different approach, incorporating four, basic statistics (which are contextualized on a play by play basis) regarding rushing, passing, scoring and play success. However, even though their approach did match 16 of the committee's 24, top four teams over the last six years, in 2015, they ranked the #3 seed Michigan State as #23, and, Mississippi (CFP ranking #12) as the #3 team, as well as LSU (#20 according to the final CFP ranking) as the #3 team in 2016 – just to mention a few significant outliers (from the committee's choices). As a byproduct of applying their model, they have also generated probabilities regarding the likelihood that certain teams will be selected into the top four;

however, the outliers listed above don't induce much confidence in said likelihoods. (<https://massey-peabody.com/college-football-2016-weekly-rankings/> is where the ratings can be found for 2016, and changing that embedded year retrieves other year's final ratings; each year can also be accessed directly from the *Archives* heading on this group's primary web page.)

The power rating system (Carroll et al, 1988), when incorporating the actual MOV, matched 16 of the 24 teams selected by the CFP committee, while this same system, when ignoring MOV, matched 20 of those same 24 teams. (The ESPN *strength of schedule* metric has had roughly the same success, when *predicting* which teams will be invited to compete in the CFP, as when MOV is ignored when calculating every team's power rating.) Another system which matched 21 of the 24 top four teams is the Comparative Performance Index (CPI), which is a straightforward calculation that is somewhat similar to the original Rating Percentage Index (RPI), though the CPI is nonlinear in format. "CPI Rating = $W\%^3 \times Ow\%^2 \times Oow\%^1$, where $W\%$ is the team's win percentage, $Ow\%$ is the team's opponent's win percentage *independent of the team*, and $Oow\%$ is the team's opponents' opponent's win percentage *independent of the teams' opponents*." (This quote can be found on www.cpiratings.com, which also provides access to weekly CPI ratings; results concerning the CPI rating formula appears later on.)

All of the above strategies in this section have tried to determine who the *best* four teams are, applying different criterion and techniques, and all of them have had from moderate to quite reasonable success with regards to matching the committee's final, top four selections. There appears to be only two published articles (Trono, 2016 & 2019) where attempting to devise a particular methodology to objectively match the exhibited behavior manifest in the final selections by the CFP committee – of its top four teams – is the main focus (rather than describing one more strategy to determine who the *best* teams are). The two WL models in the latter article – both with and without MOV – have now also matched 21 of the 24, top four teams over the first six years of the CFP. (The ultimate goal would be to discover a strategy that reproduces the same two semi-final football games, as announced by the CFP committee, *after* the final weekend of the NCAA football season.)

A similar situation occurs every spring when the NCAA men's basketball tournament committee decides which teams – besides those conference champions who are awarded an automatic bid – will receive the remaining, at-large invitations to the NCAA men's basketball tournament. Several articles have described particular models that project who this committee will invite, based upon the teams that previous committees have selected (Coleman et al, 2001 & 2010).

3. The Initial Linear Model

As stated previously, the power rating system, both with – and without – MOV is a reasonable predictor of the committee's top four teams selected: when excluding MOV, 20 of the 24 top four teams selected appear in this power rating's top four, from 2014-2019, and six teams appear in the exact same, ranked position that the committee chose; when including MOV, there were seven exact matches and 16 teams were correctly chosen. The simplest, linear combination of these two ratings, utilizing weights of +1, would generate seven exact matches and 17 selections that agree with the committee's choices.

In a manner similar to Coleman et al (2001 & 2010), the first four years of the CFP committee's final, top four team selections were used as training data to determine which weights would be the most accurate in a linear equation that initially included just three team attributes:

the team's power rating when MOV is ignored, the power rating when the full MOV is included, and the number of losses for the team that year. Games where FBS teams played against teams which are not in the FBS incorporate one generic team name (e.g. NON_DIV1A) that represents all of those non-FBS teams, for the purpose of calculating the non-MOV ratings; those games are omitted when MOV is involved (during the rating calculations) to avoid blowout wins over weak teams overly influencing said ratings.

Monte Carlo techniques led to the discovery of many sets of weights that matched 14 of the 16 teams selected from 2014 to 2017 (with nine team ranks being identical to the committee's). Therefore, to select the best performing weights, from amongst those many possible candidates, the weights which produced the highest average Spearman Correlation Coefficient (SCC) values across the top 25 (with nine exact matches, and 14 correct selections overall), for those four years, would be chosen from the one million, randomly generated sets of weights after incorporating one somewhat subtle observation.

When generating/evaluating the first one million sets of random weights, it appeared that those weights which produced the highest accuracy were not uniformly distributed throughout the pseudorandom number generator's range (from zero to one). Since the difference between two team's power ratings, when MOV was included, is typically significantly much larger than when MOV was excluded, the same weight multiplying the MOV-based power rating created a larger overall separation between two teams than when that same weight multiplied the team's non-MOV power ratings instead. So, three random values (between zero and one) were generated, but the random weight to be paired with the non-MOV power rating was multiplied by 100, and the random value to multiply the number of losses was increased by a factor of ten (and this weight, when multiplied by the team's number of losses, is subtracted from the other two products). This increased the number of weights which achieved the best performance (nine exact matches, and 14 overall) from 11 to 5,119 (out of one million random sets of weights).

4. The Improved Linear Model

When examining the power ratings that were calculated at the end of the 2016 season (each of which is the sum of the difference between the average offensive and defensive point totals for that team, OD, plus that team's computed strength of schedule component, SOS), it is impossible for the CFP committee's #2 and #3 (both one loss) teams to appear in the same order that the committee ranked them since the two computed values for #3 Ohio State are both larger than those for #2 Clemson: power ratings of 37.5 vs. 24.24 with MOV, and 1.18 vs. 1.06 without.

In 2017, the undefeated Central Florida (UCF) team was ranked #12 by the committee, however, the initial linear model considered them to be the #4 team. Given the two relatively low SOS values for UCF, perhaps a more accurate linear model could be discovered if the two power ratings were separated into their constituent OD and SOS values. Therefore, this new, improved linear model has five quantities that, when multiplied by some specific weights, would produce the value by which all teams could be ordered (to generate that year's top four teams)

With this modified model, it would then be theoretically possible for #2 Clemson to be ranked ahead of #3 Ohio State, when using the scores from the 2016 regular season; perhaps UCF might also disappear from the top four teams produced by this improved linear model (in 2017), after examining the results when applying the most accurate, five random weights discovered (instead of the three for the initial linear model).

5. Results

When applying the Monte Carlo approach to this improved linear model, which now utilizes five weights, there were once again many more sets of weights generated that matched the committee’s top four choices when the no-MOV weights were first multiplied by 100, and the punitive weight, associated with each team’s number of losses, was multiplied by ten. The highest average, top 25 SCC value, with 14 of the 16 teams being matched – and nine teams in the exact position as chosen by the committee (over the one million random weight sets), was somewhat higher (0.8392308 versus 0.8177884) in the new, improved linear model than when the power rating wasn’t separated into its two constituent components. Therefore, this updated linear prediction model was chosen as the one to assess against the 2018 and 2019 seasons.

Of course there is no guarantee that subsequent years will be as predictable as 2018, but the accuracy of the improved linear model, after training with the first four final rankings chosen by the CFP, is quite exemplary. Appendix A contrasts the top eleven teams in the final CFP committee ranking, from 2014 to 2019, with where the improved linear model ranked them; one can see that, not only do the CFP committee’s top four teams in 2018 appear in the correct positions, but also the next four teams matched exactly the committee’s ranking as well. In 2019, the final four teams were also correctly selected by this model, though the top two teams produced by in the improved linear model are reversed from the ordering released by the committee. The five weights that were discovered during the Monte Carlo process are: full MOV OD weight = 0.30912775; full MOV SOS weight = 0.83784781; no-MOV OD weight = 85.99451009; no-MOV SOS weight = 49.28798644; and a penalty per loss of 0.44385664. With these five weights, the number of exact matches is 15, and 22 of the 24 top four teams selected by this model – from 2014 to 2019 – also appear in the CFP committee’s top four those six years. (It is somewhat surprising to notice that the full MOV SOS weight is almost three times the OD weight, whereas the no-MOV SOS weight is roughly half of the no-MOV OD weight.)

The SCC values for 2016 seem to be significantly lower than the other five years, since the CFP was instituted, and that is primarily due to four teams having low power ratings as opposed to where the committee ranked them. (These large differences – between the predicted and actual positions of each team – are then squared during the SCC top 25 calculation.) Here are those four teams, with their CFP ranking, their predicted ranking (using the five parameter model), and their power rating rankings (both with – then without – MOV): Oklahoma State (12, 25, 36, 31); Utah (19, 33, 24, 33); Virginia Tech (22, 30, 17, 32); and Pittsburgh (23, 34, 29, 29).

Table 1 – SCC values when comparing results against the CFP committee’s top 25 choices.

Year	SCC_Ones	SCC_MC	SCC_Best
2014	0.5288462	0.9292308	0.9461538
2015	0.3373077	0.8546154	0.9123077
2016	0.3503846	0.7088462	0.7434615
2017	0.6423077	0.8642308	0.9030769
2018	0.4769231	0.8619231	-----
2019	0.6792308	0.8623077	-----

(All five weights were +1 for the improved linear model in the SCC_Ones column above, and the weights discovered during the Monte Carlo process produced the results in the other two columns, using different weights for each row in the SCC_Best column.)

Table 1 illustrates how much better the SCC top 25 values are for the five parameter linear model, whose weights were selected via the Monte Carlo process, over those which were derived when using five weights of '+1' (while continuing to subtract the product associated with each team's number of losses). The rightmost column contains the best, one year match that was found during the Monte Carlo evaluation process, where one million random weights were generated, and where a different set of five weights was used to produce the results in each row for the SCC_Best column. (It is unclear if there might exist another linear model, incorporating other rating systems – or team statistics – that might improve upon the results reported here. Or, if perhaps some other nonlinear/hybrid prediction model could match the committee's selections more accurately. And please remember: the goal here is *not* to attempt to determine who the *best* four teams are, but, to try and match who the committee has chosen to be the top four teams that year.)

To fairly compare the results of this improved linear model with the CPI, the training data (i.e. first four, final CFP rankings) should be excluded. Examining the 31 rankings that were announced, from 2014-2019, before the release of the final rankings, there were 60 exact matches, and 104 teams in the top four, as determined by the CFP committee (out of 124 appearing there), that were generated by the improved linear model, whereas the CPI had 36 exact matches and 97 teams correctly appearing in the top four. For the last two, final rankings, 2018 and 2019, which are outside of the training data set, the improved linear model had all eight teams, as chosen by the committee – with six exact matches, and CPI had four exact matches, and six chosen, top four teams.

6. Confidence Estimate for CFP Selection

After examining the final ratings that are produced by the improved linear model, all teams with a rating of 100 or higher have appeared in the CFP committee's top four. It also seems that any team that was being considered for an invitation had a rating of at least 80, so a strategy to estimate how likely a team would be invited, was devised. Using 80 as the threshold, all teams with ratings of at least 80 would have their ratings summed (after subtracting 80), and that total would be used to normalize each team's estimated chance to be invited. For instance, in 2014, the sum of adjusted ratings was 87.87, so TCU had an estimated chance of being designated a top four team of 39.6%: $(88.7 - 80) / 87.87 * 4$. (The final multiplication is because four teams are invited.) However, given some teams have more than 100% chance of being invited, after performing this calculation, an adjustment is made so that any excess above 100% is proportionately added to those teams being considered – whose estimate eventually remains below 100%. After the necessary adjustment, TCU's estimate became 44.1%.

Appendix B lists both the raw estimates as well as the adjusted estimates for the small number of teams that this approach considers. For the first three years, six, seven, and six teams were above the threshold, and all 19 teams were in the top six (or seven) positions in those final rankings. In 2017, the only year the improved linear model disagreed with who the committee chose, ten teams were above the threshold, and one of those teams was ranked #12 by the committee. There were seven such eligible teams in 2018 and six in 2019; one team in 2019 was not considered all that strongly by the committee that year (ranked as #17). As a first approximation, this particular methodology seems to provide a reasonable assessment of the likelihood for a team to be selected by the CFP committee to compete for the national championship.

7. Other Systems

Dr. Kenneth Massey’s web page (<http://www.masseyratings.com/cf/compare.htm>) – and the accompanying archive of previous rankings – is a repository for how each of the many computer-based systems that appear on it have ordered the teams (up to that point in the season). Table 2 lists how many times the CFP committee’s top four also appeared in the top four of these rankings (excluding the CFP ranking, and the AP and coaches’ poll which also appear on said web page as well.) Many of the systems listed on the Massey comparison page ignore MOV; other systems use the full MOV – while some systems may compress the MOV before performing any calculations. Four of the 24 CFP participants selected (since 2014) have appeared in every listed system’s top four: nine of these 24 teams appeared in the top four roughly 90% of the time; three more appear close to 80%; five at roughly 50%; one around 40% (CFP #4 Alabama, in 2017); and the other two teams – one around 20% and the other around 35% (#4 Oklahoma, in 2018 is the former, and #4 Oklahoma in 2019 is the latter).

Table 2 – Comparing the CFP committee’s top four team selections against systems listed online.

CFP	#1	#2	#3	#4
2014	Alabama	Oregon	Florida State	Ohio State
126 systems	124	118	63	76
2015	Clemson	Alabama	Michigan State	Oklahoma
125 systems	110	124	66	107
2016	Alabama	Clemson	Ohio State	Washington
126 systems	126	69	122	101
2017	Clemson	Oklahoma	Georgia	Alabama
116 systems	111	62	101	47
2018	Alabama	Clemson	Notre Dame	Oklahoma
110 systems	110	110	79	24
2019	LSU	Ohio State	Clemson	Oklahoma
103 systems	102	103	101	36

Not all systems on the Massey ratings page appear there each of those six years, however, roughly 80 to 90 systems do. Table 3 highlights that six of these would also have matched at least 22 of the 24 CFP selected teams, if said system’s top four was examined. (The first six columns in Table 3 are labelled according to the three letter acronyms for these systems – as appearing on Massey’s web page.)

Table 3 – Six most accurate, online systems (as well as the others mentioned in this article).

System	BIL	DES	MvG	BSS	HAT	PPP	Imp. Lin.	CPI	Power One	Power MOV	WL One	WL MOV
Exact	16	19	8	13	8	10	15	14	6	7	10	9
Switch	2	2	6	4	6	4	4	4	2	0	2	0
Sum	18	21	14	17	14	14	19	18	8	7	12	9
Match	24	23	23	22	22	22	22	21	20	16	21	21

(The results in Table 3 are when matching the CFP committee's final, top four selections. "One" above denotes when MOV is ignored, i.e. limited to one point, and MOV is when the actual margin of victory for each game is used.)

Richard Billingsley's system (BIL) relies on each team's final rating from the year before to seed the next year's rating, whereas most systems incorporate only the scores from that year to determine each team's rating that year. (BIL was one of the six BCS computer-based systems.) Some of the six systems listed in Table 3 are transparent in how their ratings are computed; the description for several of them are not available online.

The row in Table 3 labelled *Switch* totals the number of teams whose ranking wasn't the same as the committee's, but said ranking didn't produce a CFP pairing any different from that agreed to by the CFP committee, i.e. the #2 and #3 teams, or the #1 and #4 teams, could be switched, or, those two pairs could be switched with the other pair of teams, and the semi-final matchups would still remain the same (as the committee created). Two of the six systems in Table 3 (that appear on Massey's page), and the CPI, ranked the top four in 2014 as Florida State, Alabama, Ohio State, and finally Oregon, so Florida State and Oregon would still play each other, as would Alabama and Ohio State; the committee's ranking of Alabama, Oregon, Florida State and Ohio State produced the same matchups. One more of those six systems, and the improved linear model, would also have chosen the same semifinal matchups in 2014, having only reversed the committee's #2 and #3 ranked teams. (The improved linear model would have 19 – out of its 22 exact correct, top four selections – appear in the *Sum* row as well; only the DES system had a larger value in the *Sum* row, which is simply the result of adding each model's number of exact positional matches to its determined *Switch* value.)

The BIL system matched the committee exactly in 2015, 2018 and 2019, while the DES system did so in 2014 and 2015, and the BSS system in 2018 and 2019. (The improved linear model matched exactly in 2015 and 2018.) Across all the systems listed on these online ratings pages, once the regular season – including all conference championship games – had concluded, the triples that follow include the number of non-poll-based orderings on Massey's web page that year, the number of systems that correctly had the same top four teams (possibly in a different order from the committee), and the number that matched the committee's top four ranking exactly that year: 2014 (123, 34, 4), 2015 (122, 41, 4), 2016 (123, 43, 5), 2017 (113, 10, 0), 2018 (107, 10, 4), and 2019 (103, 35, 5).

8. Overfitting the Improved Linear Model

With regards to the Monte Carlo approach that was used with the improved linear model, one set of weights did match ten committee selections exactly (instead of the nine exact matches that occurred with the aforementioned weights for the improved linear model); however, only three other non-exact matches occurred – for a total of 13 teams matched (instead of 14). The average, top 25 SCC value was also quite a bit lower (0.77285) than the values for the chosen weights (0.83923), which – along with the lower, overall match total – is why this other set of weights was rejected.

If this other set of weights *had* been chosen, because of the ten exact matches, then perhaps, in hindsight, it might have been claimed that these weights had possibly been over-fitted because when applied to the 2018 season, only one additional exact match would've been generated (and three overall matches, within the committee's final ranking), and likewise in 2019, whereas the

improved linear model matched all top four teams exactly in 2018, and simply reversed the CFP committee's top two teams in 2019. When considering all six CFP years (2014-2019), a total of 16 exact matches, and 19 overall, were generated for this possibly over-fitted set of weights, where the latter is three fewer than the chosen weights for the improved linear model (which had 22 correct teams *chosen* into the top four).

9. Conference Playoff Champion Bonus

In a manner similar to Coleman et al (2010), when additional parameters were added to their previous model (2001) – to investigate any possible conference bias during the selection of the at large teams for the NCAA, men's basketball tournament, another parameter could be included in the improved linear model described here, along with another weight to multiply it by: did this team win its conference championship game? (If such a parameter were to be added, should this additional parameter only apply to the major conferences, or to all of them?) This additional parameter was ultimately left out of the improved linear model since several teams who did not even *play* in their conference's championship game *were* invited to compete in the CFP: #4 Oklahoma, in 2015; #3 Ohio State, in 2016; #4 Alabama, in 2017; and #3 Notre Dame, in 2018. (And please remember: the purpose of the improved linear model is to match the CFP committee's selections – and *not* to try and determine which teams *should* be in the top four/CFP, though hopefully those two objectives should be synonymous.)

10. Summary

While the CFP committee may never be replaced by a completely objective methodology, the analysis presented here indicates that the committee's selections have been in reasonable agreement with many objective ranking strategies. It seems that the committee's first four years (2014-2017) of creating its list of the top 25 teams has been captured fairly well in the improved linear model described here – which matched the top eight teams exactly (in the committee's final ranking) in 2018, and has also agreed with both of the semifinal matchups, as determined by the CFP committee, in four of the six years that the CFP has been active (only failing to do so in 2017 and 2019). And even though the training set was quite small (the first four, final rankings), when applied to the 33 other rankings established by the CFP committee, the performance of the improved linear model remained consistently accurate.

When compared against the CPI, the improved linear model has correctly placed teams into the top four seven more times (104 to 97) in those *non-binding* rankings as well as matching one more team of the six, final top four rankings from 2014-2019: 22 for the improved linear model, and 21 for CPI. A complete, weekly breakdown of the improved linear model's rankings can be found at <http://jtrono.knightsites.co/improved-linear-model/>. Besides the accurate performance with regards to the CFP, the improved linear model would also have matched 26 of the 32 top two teams during the BCS era (1998-2013).

The average difference between the improved linear model's rankings and the last week of CFP committee rankings is 2.8667, for the past six years, with a corresponding SCC value of 0.84690 (for those 25 teams). The average difference between the two post bowl polls and the final, improved linear model ranking is 2.5, with SCC values of 0.89548 for the top 25 teams in the writers' poll as well as 0.88755 for the coaches' poll. (Results from 1998-2013 are quite similar to those just listed.)

At this time, the improved linear model described here now surpasses the CPI as the most accurate model (when matching the top four team decisions made by the CFP committee), whose specific methodology for determining those teams has been made public. Perhaps, the publication of this benchmark will allow for further model development in pursuit of an even more accurate model.

11. Acknowledgements

A very sincere thank you is given to the anonymous reviewer for pointing out several shortcomings of a preliminary submission of this work. Those comments contributed quite significantly to this work, especially with regards to the depth – and the expanded scope - of the analysis contained herein.

References

- Carroll, B., Palmer, P., Thorn, J., 1988, *The Hidden Game of Football*, Warner Books.
- Coleman, B. J., Lynch, A. K., 2001, Identifying the NCAA Tournament ‘Dance Card’, *Interfaces* 31(3):76-86.
- Coleman, B. J., Dumond, J. M., and Lynch, A. K., 2010, Evidence of Bias in NCAA Tournament Selection and Seeding, *Managerial and Decision Economics* 31:431-452.
- Trono, J., 2016, Is it Possible to Objectively Generate the Rankings Produced by the College Football Playoff Committee, SMC Tech Report SMC-2016-CS-001, 2016, http://academics.smcvt.edu/jtrono/Papers/CFP_And_Partition.pdf.
- Trono, J., 2019, Objectively Modelling the College Football Playoff Committee’s Selections, Proceedings of the 7th International Conference on Mathematics in Sport, 2019, SMC Tech Report SMC-2019-CS-001, http://academics.smcvt.edu/jtrono/Papers/CFP_Five_Years.pdf.

Appendix A – Results for the Improved Linear Model: CFP ranking from #1 through #11.

CFP Rank	2014	Imp.Lin. model	W	L	No-MOV		Full MOV	
					OD	SOS	OD	SOS
1	Alabama	104.43	12	1	0.85	0.29	19.33	13.74
3	Florida State	103.96	13	0	1.00	0.19	10.67	6.33
2	Oregon	98.76	12	1	0.85	0.28	21.75	6.66
4	Ohio State	90.38	12	1	0.85	0.17	24.08	2.28
6 (#5)	TCU	88.70	11	1	0.83	0.09	25.82	6.39
5	Baylor	81.64	11	1	0.83	0.01	21.09	4.41
7	MississippiSt	78.53	10	2	0.50	0.36	16.73	9.85
9	Mississippi	75.38	9	3	0.67	0.17	13.73	13.99
14	UCLA	74.66	9	3	0.50	0.46	5.42	10.32
10 (#10)	Arizona	69.31	10	3	0.54	0.34	7.38	6.16
8	MichiganState	68.45	10	2	0.67	0.08	21.82	1.23
11	Kansas State	64.22	9	3	0.50	0.20	11.82	10.79

CFP Rank	2015	Imp.Lin. model	W	L	No-MOV		Full MOV	
					OD	SOS	OD	SOS
1	Clemson	102.18	13	0	1.00	0.08	16.50	8.53
2	Alabama	97.90	12	1	0.85	0.27	17.17	7.92
3	MichiganState	96.54	12	1	0.85	0.32	11.62	5.40
4	Oklahoma	94.12	11	1	0.83	0.20	25.00	6.69
7 (#5)	Ohio State	88.89	11	1	0.83	0.20	21.00	1.92

5	Iowa	83.89	12	1	0.85	0.11	13.33	2.03
6	Stanford	82.36	11	2	0.69	0.31	14.15	5.08
8	Notre Dame	78.73	10	2	0.67	0.23	12.33	8.18
13	Northwestern	72.41	10	2	0.67	0.22	0.91	5.44
11 (#10)	TCU	70.93	10	2	0.67	0.10	11.27	6.91
18	Houston	68.85	12	1	0.95	-0.14	19.50	-3.50
16	OklahomaState	67.79	10	2	0.67	0.07	11.09	6.19
10	NorthCarolina	67.76	11	2	0.69	-0.03	15.64	7.11
9	Florida State	66.08	10	2	0.67	0.01	14.55	5.21

CFP		Imp.Lin.			No-MOV		Full MOV	
Rank	2016	model	W	L	OD	SOS	OD	SOS
1	Alabama	114.63	13	0	1.00	0.25	28.83	8.83
3	Ohio State	104.49	11	1	0.83	0.35	28.50	8.95
2	Clemson	93.94	12	1	0.85	0.22	18.67	5.58
4	Washington	93.37	12	1	0.85	0.15	26.33	6.21
6 (#5)	Michigan	87.16	10	2	0.67	0.30	28.50	8.23
5	Penn State	84.49	11	2	0.69	0.29	13.31	9.18
10	Colorado	71.28	10	3	0.54	0.26	9.33	12.52
8	Wisconsin	71.27	10	3	0.54	0.29	13.15	9.32
15	West.Michigan	69.63	13	0	1.00	-0.31	22.00	-9.41
7 (#10)	Oklahoma	68.38	10	2	0.67	0.08	15.00	3.67
9	Southern Cal	66.92	9	3	0.50	0.27	10.75	10.29
11	Florida State	61.92	9	3	0.50	0.24	7.82	7.17

CFP		Imp.Lin.			No-MOV		Full MOV	
Rank	2017	model	W	L	OD	SOS	OD	SOS
1	Clemson	102.88	12	1	0.85	0.31	19.67	10.58
3	Georgia	98.31	12	1	0.85	0.25	21.17	8.11
6	Wisconsin	95.44	12	1	0.85	0.23	20.62	6.06
12	Cent.Florida	93.04	12	0	1.00	0.04	22.82	-2.36
5(#5)	Ohio State	90.07	11	2	0.69	0.36	22.54	8.25
4	Alabama	89.11	11	1	0.83	0.14	25.00	4.24
2	Oklahoma	87.16	12	1	0.85	0.05	19.92	7.03
9	Penn State	84.37	10	2	0.67	0.30	26.08	5.72
10	Miami(F.)	80.53	10	2	0.67	0.25	7.91	10.78
8 (#10)	So.Cal	80.10	11	2	0.69	0.28	8.23	6.33
14	Notre Dame	78.41	9	3	0.50	0.49	13.50	10.05
7	Auburn	71.38	10	3	0.54	0.26	17.33	9.67
11	Washington	71.03	10	2	0.67	0.12	19.36	2.87

CFP		Imp.Lin.			No-MOV		Full MOV	
Rank	2018	model	W	L	OD	SOS	OD	SOS
1	Alabama	133.02	13	0	1.00	0.20	33.08	8.29
2	Clemson	110.59	13	0	1.00	0.25	30.92	3.24
3	Notre Dame	105.57	12	0	1.00	0.25	16.50	2.57
4	Oklahoma	91.71	12	1	0.85	0.15	17.08	7.62
5 (#5)	Georgia	91.58	11	2	0.69	0.34	18.58	12.69
6	Ohio State	89.12	12	1	0.85	0.15	17.77	4.28
7	Michigan	80.29	10	2	0.67	0.27	19.25	5.13
8	Cent.Florida	79.87	12	0	1.00	-0.16	21.55	-5.85
10	Florida	64.37	9	3	0.50	0.20	6.90	12.79
11 (#10)	LSU	63.78	9	3	0.50	0.21	9.00	10.72
14	Kentucky	62.22	9	3	0.50	0.20	7.82	9.88

13	Washington St	59.83	10	2	0.67	-0.03	13.45	0.50
12	Penn State	59.17	9	3	0.50	0.19	14.58	4.34
9	Washington	58.43	10	3	0.54	0.13	8.58	4.96

CFP		Imp.Lin.			No-MOV		Full MOV	
Rank	2019	model	W	L	OD	SOS	OD	SOS
2	Ohio State	121.39	13	0	1.00	0.36	36.15	7.73
1	LSU	110.15	13	0	1.00	0.17	24.58	9.76
3	Clemson	94.38	13	0	1.00	-0.02	35.17	-1.79
4	Oklahoma	90.24	12	1	0.85	0.09	15.58	9.95
17 (#5)	Memphis	82.19	12	1	0.85	0.09	14.92	0.58
5	Georgia	80.39	11	2	0.69	0.20	16.42	8.36
10	Penn State	79.39	10	2	0.67	0.23	15.55	7.78
6	Oregon	77.02	11	2	0.69	0.14	19.25	6.83
8	Wisconsin	76.42	10	3	0.54	0.35	18.54	9.94
15 (#10)	NotreDame	74.00	10	2	0.67	0.21	18.42	1.47
9	Florida	73.30	10	2	0.67	0.09	14.00	9.32
11	Utah	72.60	11	2	0.69	0.08	19.92	4.83
7	Baylor	70.55	11	2	0.69	0.03	13.92	7.54

Appendix B – Estimated confidence (Conf.) for selection (to CFP)

CFP		Imp.Lin.			Conf. Adjusted	
Rank	2014	model	W	L	Conf.	Adjusted
1	Alabama	104.43	12	1	111.2	100.0
3	Florida State	103.96	13	0	109.1	100.0
2	Oregon	98.76	12	1	85.4	95.0
4	Ohio State	90.38	12	1	47.2	52.6
6 (#5)	TCU	88.70	11	1	39.6	44.1
5	Baylor	81.64	11	1	7.5	8.3

CFP		Imp.Lin.			Conf. Adjusted	
Rank	2015	model	W	L	Conf.	Adjusted
1	Clemson	102.18	13	0	103.3	100.0
2	Alabama	97.90	12	1	83.4	84.3
3	MichiganState	96.54	12	1	77.0	77.9
4	Oklahoma	94.12	11	1	65.8	66.5
7 (#5)	Ohio State	88.89	11	1	41.4	41.9
5	Iowa	83.89	12	1	18.1	18.3
6	Stanford	82.36	11	2	11.0	11.1

CFP		Imp.Lin.			Conf. Adjusted	
Rank	2016	model	W	L	Conf.	Adjusted
1	Alabama	114.63	13	0	141.1	100.0
3	Ohio State	104.49	11	1	99.9	100.0
2	Clemson	93.94	12	1	56.9	71.6
4	Washington	93.37	12	1	54.5	68.6
6 (#5)	Michigan	87.16	10	2	29.2	36.8
5	Penn State	84.49	11	2	18.3	23.0

CFP		Imp.Lin.			Conf.
Rank	2017	model	W	L	Conf.
1	Clemson	102.88	12	1	90.6

3	Georgia	98.31	12	1	72.5
6	Wisconsin	95.44	12	1	61.1
12	Cent.Florida	93.04	12	0	51.6
5(#5)	Ohio State	90.07	11	2	39.9
4	Alabama	89.11	11	1	36.1
2	Oklahoma	87.16	12	1	28.4
9	Penn State	84.37	10	2	17.3
10	Miami(F.)	80.53	10	2	2.1
8 (#10)	So.Cal	80.10	11	2	0.4

CFP		Imp.Lin.				
Rank	2018	model	W	L	Conf.	Adjusted
1	Alabama	133.02	13	0	149.5	100.0
2	Clemson	110.59	13	0	86.2	100.0
3	Notre Dame	105.57	12	0	72.1	87.8
4	Oklahoma	91.71	12	1	33.0	40.2
5 (#5)	Georgia	91.58	11	2	32.6	39.7
6	Ohio State	89.12	12	1	25.7	31.3
7	Michigan	80.29	10	2	0.8	1.0

CFP		Imp.Lin.				
Rank	2019	model	W	L	Conf.	Adjusted
2	Ohio State	121.39	13	0	167.6	100.0
1	LSU	110.15	13	0	122.1	100.0
3	Clemson	94.38	13	0	58.2	100.0
4	Oklahoma	90.24	12	1	41.5	79.6
17 (#5)	Memphis	82.19	12	1	9.0	17.3
5	Georgia	80.39	11	2	1.6	3.1