Statistical Insights into Florida State's Controversial Omission from the 2023 College Football Playoff

Written by: Ben Katzman

Introduction:

On December 3rd, 2023, the college football playoff (CFP) committee sent shockwaves through the sports world after announcing their decision to leave Florida State out of their four team playoff. This choice struck many as odd, as Florida State had just days before completed an undefeated season and had been crowned ACC champion by knocking off their rival Louisville, the 14th ranked team in the country. Despite this, the committee elected to invite Alabama and Texas, who, while both conference champions (SEC and Pac-12, respectively), both had a loss on their record.

Outraged, fans let their voices be heard, accusing the committee of bias towards the more popular teams instead of selecting the most deserving squad. Even the Florida State athletic director, Michael Alford, spoke up on this debate, saying, "The argument of whether a team is the 'most deserving OR best' is a false equivalence. It renders the season up to yesterday irrelevant and significantly damages the legitimacy of the CFP. The 2023 Florida State Seminoles are the epitome of a total TEAM. To eliminate them from a chance to compete for a national championship is an unwarranted injustice that shows complete disregard and disrespect for their performance and accomplishments. It is unforgivable." (Gaydos)

But was the CFP committee justified in their decision to exclude Florida State? Was there any merit behind such a wildly unpopular choice? In this paper, we will be investigating different statistical methods that can be used when attempting to rank teams and different factors that play into how they are graded. We will then be producing our own top 25 teams, comparing and contrasting them to those produced by the CFP committee. Finally, we will be comparing Florida State to a group of teams of similar merit to see if we can definitively say that they are more deserving of a spot in the 2023 CFP.

Methodology:

The data set we will be using to carry out this statistical analysis is a win-loss matrix describing all 2023 games among Division 1A teams up through week 12 (including week 0) and conference championship games. This matrix has a row for each contest, a home win column indicating whether the winner was the home (1) or away team (0), and a column for each Division 1A team. Entries in a team's column show as 0 for games that the team didn't play in, 1 for the team's home games, and -1 for the team's away games.

For the purposes of this analysis, we will be using a Bradley-Terry model to fit the matchup data. The Bradley-Terry model is a popular statistical method used for analyzing pairwise comparison data, commonly applied in sports analytics to assess the relative abilities of competing teams. The model assigns a strength to every team based on the observed outcomes of their matchups against opponents. We are required to assign a reference team that all others will be compared to, given a strength of 0. For simplicity, we will be using UMass as the reference team, due to the fact that they were a poor team in 2023 (a record of 3-9) and thus had a low chance of getting ranked in the top 25.

The reason we will be using Bradley-Terry modeling is because it offers key advantages over less sophisticated measures. One such advantage is its ability to account for the varying strengths or abilities of competitors in a more nuanced manner. Unlike win-loss percentages that oversimplify outcomes or sum scores that ignore the quality of opponents faced, Bradley-Terry models consider the relative strengths of opponents. This means that the model incorporates the probability of winning or losing based on the relative skills of competitors, providing a more accurate assessment of performance.

The first step in this analysis was to fit a regular Bradley-Terry model and a Badley-Terry model that accounts for game location. The latter is done by incorporating a parameter permitting an increased likelihood of winning for whichever of the two teams is playing a given game at home. However, since the conference championship games are played at neutral sites, but still list a home and away team, they will be excluded from the data for this analysis. We will then compare the two models, using the Akaike Information Criterion, to find whether incorporating a game location parameter makes the model fit better.

We will once again fit two more Bradley-Terry models, this time using ordinary logistic regression and Bayesian logistic regression. The key difference between the two lies in their approach to estimating model parameters and handling uncertainty. For the ordinary logistic regression model, the goal is to find the parameter values that maximize the likelihood function, essentially finding the most probable values given the data. Alternatively, the Bayesian logistic model treats the parameters as random variables and updates their probability distributions based on observed data, producing posterior distributions that reflect both the prior information and the current data. After this, we are left with four distinct Bradley-Terry models that we must choose between to proceed. To make this decision, we will run 5 fold cross-validation on each of the models in order to estimate their generalization error.

Using the best model we find from this analysis, we will then output our own top 25 teams in college football according to their Bradley-Terry strength index. This allows us to compare our findings to that of the CFP committee and see how different the results are.

Finally, we will compare Florida State to three other teams that they were directly competing against for a playoff spot to see if they truly deserved to be in or if the committee got it right. The first way that we will do this is simply by looking at the Bradley-Terry strength indexes output by our best model and comparing to see who's the highest. The second method we will use is a t-tests based on comparing the estimated difference in teams' Bradley-Terry coefficients to an associated standard error, incorporating Bonferroni as a multiplicity correction. The t-test will be two sided and we will be using a significance level of 0.05.

Results:

The first step in our analysis was to compare the normal Bradley-Terry model and the Bradley-Terry model that incorporates a parameter for home field advantage. To do this, we used the Akaike Information Criterion, resulting in values of 1724.23 for the normal Bradley-Terry model and 856.09 for the Bradley-Terry model that factors in home field advantage. This indicates the model with home field advantage offers a better representation of the underlying data when comparing the two models.

We then fit two more Bradley-Terry models, this time using ordinary logistic regression and Bayesian logistic regression. We now have four distinct models, and must choose which is best as we proceed forward. This is done by running cross-validation on each of the models with 5 folds in order to estimate the generalization error of each. Below are the results of said analysis:

Regular Bradley-Terry Model MSE	Bradley-Terry Model with Home Field Advantage MSE	Ordinary Logistic Regression Bradley-Terry Model MSE	Penalized Logistic Regression Bradley-Terry Model MSE
0.624	0.201	0.217	0.212

The above are estimated generalization errors for each version of the Bradley-Terry model calculated from cross-validation with 5 folds.

This result shows that the Bradley-Terry models with home field advantage, ordinary logistic regression, and Bayesian logistic regression all perform very similarly in terms of estimated generalization error. However, since the Bradley-Terry model with home field advantage excludes some of the data and has a slightly higher estimated generalization error, we will be proceeding forward with the Bradley-Terry models using ordinary logistic regression and Bayesian logistic regression.

Using the Bradley-Terry models using ordinary logistic regression and Bayesian logistic regression, we output our own top 25 teams in college football according to their strength index. We added the CFP committee's rankings as well so that we could compare the results. Below are the top 25's for each:

Rank	Ordinary Logistic Regression	Penalized Logistic Regression	College Football Playoff Committee
1	Michigan	Michigan	Michigan
2	Washington	Washington	Washington
3	Ohio State	Florida State	Texas
4	Florida State	Ohio State	Alabama
5	Oregon	Alabama	Florida State
6	Penn State	Georgia	Georgia
7	Liberty	Oregon	Ohio State
8	Alabama	Texas	Oregon
9	Georgia	Penn State	Missouri
10	Texas	Ole Miss	Penn State
11	Ole Miss	Liberty	Ole Miss
12	LSU	LSU	Oklahoma
13	Missouri	Missouri	LSU
14	Oklahoma	Oklahoma	Arizona
15	Iowa	Louisville	Louisville
16	Utah	Iowa	Notre Dame
17	Louisville	Notre Dame	Iowa
18	Notre Dame	Arizona	NC State
19	Arizona	Utah	Oregon State
20	Oregon State	Oregon State	Oklahoma State
21	Tennessee	James Madison	Tennessee
22	NC State	NC State	Clemson
23	USC	Tennessee	Liberty
24	Kansas State	Kansas State	SMU
25	Clemson	USC	Kansas State

The above chart shows the top 25 teams in college football according to the Bradley-Terry strength index from the ordinary logistic regression and Bayesian logistic regression models. The far right is the actual CFP committee's rankings from December 3rd, 2023.

As we can see, fitting the model with a penalty does change the teams' relative scores, as the rankings differ between the two methods. When comparing the ordinary logistic regression model's top 25 rankings to that of the CFP committee, we can clearly see that Florida State is ranked above both Alabama and Texas in our model but not in the CFP rankings. This shows our model projects Florida State in the top 4 and thus in the CFP.

The last step of our analysis is to compare Florida State's case against those of Alabama and Texas (who the playoff committee put in over Florida State), as well as Ohio State. We have chosen Ohio State as the third team due to the fact that our ordinary logistic regression Bradley-Terry model puts them over Florida State, despite Ohio State not being a conference champion and having one loss on their record. We will first be comparing the Bradley-Terry index of each team according to the Bayesian logistic regression model. We will be using the ordinary logistic regression Bradley-Terry model for this analysis since the estimated generalization error is the same, so we choose the less complex model. This results in a team strength of (in descending order) 37.42 for Ohio State, 24.30 for Florida State, 9.27 for Alabama, and 8.05 for Texas.

Lastly, we will run a t-tests based to compare the estimated difference in teams' Bradley-Terry coefficients to an associated standard error, incorporating Bonferroni as a multiplicity correction. This t-test results in an estimated difference in Bradley-Terry strength from Florida State of -13.12 for Ohio State, 15.02 for Alabama, and 16.25 for Texas. However, each of these values have an incredibly high p-value of 1.00 for all of the values.

Discussion:

With all of this information, let's get to the question pressing on everybody's mind: does Florida State deserve to make the CFP? First, let's start with the model selection. We first make the decision to proceed with the Bradley-Terry model that factors in home field advantage over the regular Bradley-Terry model because the latter has a higher AIC. While this logic is sound, the use of the home field advantage model brings forth some concerns. Most importantly, all of the championship games were played at neutral sites, meaning there was no home team, despite there being one listed. Due to this fact, the second dataset will be smaller, excluding high leverage games involving the top teams in college football, and thus we will proceed with caution when considering this model.

Next we add the ordinary logistic regression and Bayesian logistic regression models to the mix. When picking between these four models, we get very similar cross-validation estimates for generalization error for all models except for the regular Bradley-Terry model. We will first eliminate the home field advantage model due to the factors listed above. Next, due to pretty much identical estimates for generalization error for both logistic regression models, we will be proceeding forward with the ordinary logistic regression model because it is simpler than the Bayesian logistic regression model.

Using this model, we get the Bradley-Terry strengths for the four best candidates for the last two playoff spots, resulting in Ohio State and Florida State scoring the highest. However, when conducting a t-test with Bonferroni multiplicity correction, we get incredibly high p-values of 1.00 for all of our calculations, meaning there is no significant difference between the strength of any of these teams. Because of this, we cannot definitively prove that Florida State is better than Ohio State, Alabama, or Texas, meaning the CFP committee is justified in picking any of these other three teams. However, if we were forced to pick teams for the final two playoff spots, Florida State would be one of them due to the higher Bradley-Terry strength.

Lastly, I would be remiss if I did not mention the factors that the committee takes into consideration that the models do not. For starters, the Bradley-Terry model only considers a binary outlook on a game, whether a team won or lost. However, the committee will watch a game and consider the degree in which a team won the game. A team blowing out another team holds more weight then simply squeaking by them in a close one. Another factor considered by the committee is how a team looks as they near the end of the season. While the Bradley-Terry gives equal weight to all games, the committee often gives a higher weight to games that happen towards the end of the season. This is particularly relevant when discussing injuries, as a banged up roster may begin to fall off near the end of the season. Both of these factors may help to explain the committee's decision to leave Florida State out of the playoffs. Late in the season, Florida State's star quarterback Jordan Travis suffered a severe leg injury, causing him to miss the rest of the season (Ferrante). Because of this, Florida State, despite winning the rest of their games, did not play up to the high standard they set earlier in the year. Thus, while the model may indicate that Florida State deserved to make the playoffs, it considers a narrower view on the season than the committee, causing them to come to different conclusions.

Works Cited

Ferrante, Bob. “Florida State quarterback Jordan Travis says leg injury will end his season with No. 5 Seminoles.” *AP News*, 21 November 2023, https://apnews.com/article/jordan-travis-quarterback-florida-state-season-ending-injury-a71d0a1981a3410e0ac897b51b0e30a6. Accessed 7 December 2023.

Gaydos, Ryan. “Florida State AD rips CFP officials after being left out: 'The committee failed college football today.'” *Fox News*, 3 December 2023, https://www.foxnews.com/sports/florida-state-ad-rips-cfp-officials-left-out-committee-failed-college-football-today. Accessed 7 December 2023.

Unit 3 Paper Appendix, Version 2

2023-12-12

```r
library(arm)
```

```
## Loading required package: MASS
```

```
## Loading required package: Matrix
```

```
## Loading required package: lme4
```

```
##
## arm (Version 1.13-1, built: 2022-8-25)
```

```
## Working directory is C:/Users/btkat/OneDrive/Desktop/STATS485/Unit 3/Paper
```

```r
library(boot)
```

```
##
## Attaching package: 'boot'
```

```
## The following object is masked from 'package:arm':
##
##     logit
```

```r
library(multcomp)
```

```
## Loading required package: mvtnorm
```

```
##
## Attaching package: 'mvtnorm'
```

```
## The following object is masked from 'package:arm':
##
##     standardize
```

```
## Loading required package: survival
```

```
##
## Attaching package: 'survival'
```

```
## The following object is masked from 'package:boot':
##
##     aml
```

```
## Loading required package: TH.data
```

```
##
## Attaching package: 'TH.data'
```

```
## The following object is masked from 'package:MASS':
##
##     geyser
```

```r
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.0     ✔ readr     2.1.4
## ✔ forcats   1.0.0     ✔ stringr   1.5.1
## ✔ ggplot2   3.4.4     ✔ tibble    3.2.1
## ✔ lubridate 1.9.3     ✔ tidyr     1.3.0
## ✔ purrr     1.0.1
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ tidyr::expand() masks Matrix::expand()
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ✖ tidyr::pack()   masks Matrix::pack()
## ✖ dplyr::select() masks MASS::select()
## ✖ tidyr::unpack() masks Matrix::unpack()
## ℹ Use the ]8;;http://conflicted.r-lib.org/conflicted package]8;; to force all conflicts to become errors
```

```
data = read.csv("https://dept.stat.lsa.umich.edu/~bbh/s485/data/cfb_wl-2023-12-07.csv")
```

In this appendix we will be running a statistical analysis on the win-loss matrix describing all 2023 games among Division 1A teams up through week 12 (including week 0) and conference championship games. We will first be creating an ordinary Bradley-Terry model and a Bradley-Terry model that factors in home field advantage. We will then be using AIC to test which model offers a better representation of the underlying data. Next, we will fit an ordinary logistic regression and a Bayesian logistic regression. After that we will attempt to find the best model by comparing the estimations of generalization error found by running cross-validation. We will then compare the top 25 rankings found from our best model and those of the CFP committee. Lastly, we will attempt to see if Florida State is better than three other teams by comparing Bradley-Terry team strengths and by running a t-test with Bonferroni multiplicity correction.

a.

```
btm = glm(home_win ~ . - week - UMass - 1, data = data)
sort(coef(btm), decreasing=TRUE)[1:25]
```

```
##     Washington        Alabama  Florida.State          Texas       Michigan 
##      1.3307216      1.2863410      1.2422962      1.2303383      1.2149329 
##       Missouri     Ohio.State     Penn.State        Georgia         Oregon 
##      1.1721375      1.1420712      1.1365984      1.1327000      1.0920381 
##       Ole.Miss           Iowa            LSU      Tennessee     Louisville 
##      1.0918825      1.0915692      1.0781384      0.9989380      0.9704112 
##           Utah        Clemson       NC.State     Notre.Dame  Kansas.State 
##      0.9662431      0.9175166      0.9159519      0.9134485      0.9078853 
##       Oklahoma            USC        Arizona North.Carolina   Northwestern 
##      0.9065477      0.8874602      0.8717103      0.8525596      0.8408783
```

The ranking of the above teams is according to the estimates reported by the Bradley-Terry model. The higher the estimate, the higher that team’s strength. UMass is left out of the data set because they are the reference team, meaning that their estimate is 0.

b.

```
btm_ha = glm(home_win ~ . - week - UMass, data = data[data$week != 14,])
cat("AIC for Bradley-Terry Model:", AIC(btm), "\n")
```

```
## AIC for Bradley-Terry Model: 1724.23
```

```
cat("AIC for Bradley-Terry Model with Home Field Advantage:", AIC(btm_ha), "\n")
```

```
## AIC for Bradley-Terry Model with Home Field Advantage: 856.0882
```

The Bradley-Terry model with home field advantage achieves a lower AIC value than the Bradley-Terry model without home field advantage. This indicates the model with home field advantage offers a better representation of the underlying data when comparing the two.

c.

```
olr_btm = glm(home_win ~ . - week - UMass - 1, family = "binomial", data = data)
lr_btm = bayesglm(home_win ~ . - week - UMass - 1, family = "binomial", data = data)

set.seed(100)
cat("Cross-Validation Generalization Error Estimate for Bradley-Terry Model:", cv.glm(data, btm, K = 5)$delta[2],
"\n")
```

```
## Cross-Validation Generalization Error Estimate for Bradley-Terry Model: 0.6243393
```

```
cat("Cross-Validation Generalization Error Estimate for Bradley-Terry Model with Home Field Advantage:", cv.glm(d
ata[data$week != 14,], btm_ha, K = 5)$delta[2], "\n")
```

```
## Cross-Validation Generalization Error Estimate for Bradley-Terry Model with Home Field Advantage: 0.200541
```

```
cat("Cross-Validation Generalization Error Estimate for Ordinary Logistic Regression Model:", cv.glm(data, olr_bt
m, K = 5)$delta[2], "\n")
```

```
## Cross-Validation Generalization Error Estimate for Ordinary Logistic Regression Model: 0.217161
```

```
cat("Cross-Validation Generalization Error Estimate for Logistic Regression Model:", cv.glm(data, lr_btm, K = 5)$
delta[2], "\n")
```

```
## Cross-Validation Generalization Error Estimate for Logistic Regression Model: 0.2120033
```

The modeling routine that I prefer is the ordinary logistic regression model. This is because Bradley-Terry model with home field advantage uses a smaller data set, excluding high leverage games, meaning it may be less reliable. Furthermore, since the cross-validation generalization error estimate is essentially the same for both logistic regression models, we will move forward with the ordinary logistic regression Bradley-Terry model, since it is the simpler of the two choices.

d.

```
sort(coef(olr_btm), decreasing=TRUE)[1:25]
```

```
##      Michigan    Washington    Ohio.State Florida.State        Oregon
##     53.575658     39.686566     37.392584     24.266664     22.598869
##    Penn.State       Liberty       Alabama       Georgia         Texas
##     21.697276     20.649990      9.256680      8.906559      8.037464
##      Ole.Miss           LSU      Missouri      Oklahoma          Iowa
##      7.856715      7.304221      6.744888      5.704692      5.411158
##   Louisville          Utah    Notre.Dame       Arizona  Oregon.State
##      5.383648      5.354202      5.349945      5.319770      5.160703
##     Tennessee      NC.State           USC  Kansas.State       Clemson
##      4.997235      4.877138      4.817081      4.687741      4.666440
```

```
sort(coef(lr_btm), decreasing=TRUE)[1:25]
```

```
##      Michigan    Washington Florida.State    Ohio.State       Alabama
##      9.686603      9.172234      7.550173      7.256782      6.532813
##       Georgia        Oregon         Texas    Penn.State      Ole.Miss
##      6.181969      6.086925      5.521712      5.398815      5.197400
##       Liberty           LSU      Missouri      Oklahoma    Louisville
##      5.000663      4.688706      4.339854      3.606580      3.289214
##          Iowa    Notre.Dame       Arizona          Utah  Oregon.State
##      3.233965      3.232125      3.162976      3.156260      3.001558
## James.Madison      NC.State     Tennessee           USC  Kansas.State
##      2.857728      2.829439      2.814650      2.683545      2.650422
```

Fitting the model with a penalty does change the teams' relative scores, as we can see that the orders differ between the two methods. When comparing the ordinary logistic regression model's top 25 rankings to that of the CFP committee, we can clearly see that Florida State is ranked above both Alabama and Texas in our model but not in the CFP rankings.

e.
f.

```
data.frame(as.list(sort(coef(olr_btm)[c("Florida.State", "Ohio.State", "Alabama", "Texas")], decreasing=TRUE)))
```

Ohio.State <dbl>
37.39258

1 row | 1-1 of 4 columns

ii.

```
glht(olr_btm, linfct = c('Florida.State - Texas == 0',
                         'Florida.State - Ohio.State == 0',
                         'Florida.State - Alabama == 0')) %>%
summary()
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: glm(formula = home_win ~ . - week - UMass - 1, family = "binomial",
##     data = data)
##
## Linear Hypotheses:
##                                Estimate Std. Error z value Pr(>|z|)
## Florida.State - Texas == 0         16.23    2496.78   0.007        1
## Florida.State - Ohio.State == 0   -13.13    3304.72  -0.004        1
## Florida.State - Alabama == 0       15.01    2496.78   0.006        1
## (Adjusted p values reported -- single-step method)
```