Tracy Williams   December 10, 2014
Extended Abstract

<center>Preseason and Penultimate AP Polls, 2004-2013</center>

Introduction

  Most college team sports use some sort of rankings system to frame the sport. With more than 100 teams, it can be difficult to get a solid grasp on the landscape of a sport. A formalized ranking of the best can help to understand who stands where.

  With most sports, that is the extent of the power of the rankings: a helpful guide to who's good. The rankings ultimately do little in determining a champion, and often do not even aid the selection for that sport's postseason.

  However, college football has historically used its ranking systems as the final arbiter in choosing its champion. The top level of college football, the FBS, is unique in that, before 2014, it has never used any sort of playoff to determine its winners. Bowl games became the prominent form for the college football postseason. The bowls were entrenched in the college football landscape before playoffs could be devised. Lower levels of college football, without the interference of the bowls, have established 24-team playoffs. For the FBS, in the absence of a playoff, the only way for a national champion to be crowned was by an outside observer. Various organizations had their polls and awarded their champions. But how much information can these polls have? Are their biases in these polls?

  The AP poll has been awarded since 1935. It amalgamates the polls of its writers into one nationally published ranking each week of the season. Starting in 1950, it began to publish a preseason poll as well. But I started to think: do these preseason polls unfairly bias the rankings for the rest of the season? The preseason polls contain very little information compared to every ensuing poll; they are merely guesses. But if each voter votes using the previous poll as a template, then the preseason poll might make it difficult for teams to overcome their preseason rankings. Barring a loss of a team in front of them, in might be impossible to move up. And if two teams have the same record, the one that started higher will probably end higher. So I decided to test this.

Previous Research

  Once I started my investigation, I planned to just regress the preseason polls onto the final polls. However, after a quick search, I discovered that someone had already attempted this. A person named Steve Albrecht had found that the preseason poll had a significant relationship with the final polls at the 5% level. He also estimated that the preseason ranking would account for 1/20th of a spot in the final ranking. That is, if two teams had identical results against identical schedules, the team ranked 20 spots ahead in the preseason poll would be ranked 1 spot higher in the final poll.

Methodology

  It was much more difficult to find a dataset of the AP polls than I had previously imagined. Many were inconsistent in their names for teams (for example, sometimes using "USC," other times "Southern California,") and others required a subscription. As a result, I just copied the past ten years, 2004-2013, from collegepollarchive.com[1] and entered the values into an Excel spreadsheet.

  Because someone had already run regressions before, I decided to use several different approaches to analyze the polls.

  The most notable choice I made amongst all of the approaches was to focus on the penultimate poll as opposed to the final poll. I did this for several reasons.

  1) The penultimate poll determines the most important decisions. The penultimate poll determines who even gets a chance to play in the national championship, whereas the final poll just reveals the winner of the championship game. The "unfairness" of the poll is resolved before the final poll because some component actually got to be played out. Additionally, the penultimate poll determines the bowls, and the bowls determine the money

---

[1] http://collegepollarchive.com/football/ap/seasons.cfm

2) The final polls do not behave like the weekly polls. Because they behave uniquely, it feels odd to treat them as a natural successor. For example, in a normal weekly poll, if the undefeated #1 were to loss, one could expect that team to fall to somewhere in the 5-10 range. But for the final poll, that team will almost always only fall to #2.

3) The final polls have exceptional circumstances. Bowl season creates many more top 25 matchups than any normal week. As a result, there is an unnatural amount of turnover that goes into the final poll.

Because of the above, I used several methods to compare the preseason polls to the penultimate polls.

Tests/Findings

Spearman's rank correlation coefficient

Spearman's rank correlation coefficient[2] can be used to analyze statistical dependence between two sets of rankings. A positive coefficient indicates that as X ranking increases, Y also increases. A negative coefficient would indicate the converse. A coefficient of zero would indicate there is no tendency for Y to increase or decrease as X increases.

I found the Spearman's rank correlation coefficient for every year between 2004-2013.   9 out of the 10 values were negative. This would indicate that the better ranked teams in the preseason poll tend to finish lower in the penultimate poll.

In retrospect, I should have expected that. This is similar to the "Top 10/Bottom 10" scorers activity from Week 1. The best teams do not have much room for improvement and will inevitably regress. And there are more than 100 teams outside of the preseason top 25.   Some of those teams will outperform expectations and deviate towards the top of the rankings. This would result in a negative coefficient.

Following is a chart of the Spearman coefficients:

Next, I just wanted to see how many teams that were unranked in the preseason were ranked in the penultimate poll. That was a pretty straightforward count. I got the ensuing results:

I was hoping for a discernible correlation here between this chart and the Spearman coefficient chart, but there was not much.   The number of preseason unranked teams to accomplish this feat seemed to be pretty consistently around 9.

I was somewhat surprised that this number was that low. This raised my value of preseason polls. Before any games have even started, voters can tell almost 2/3 ranked teams in the final poll. Some of that relationship may be bias, but still better than I was expecting.

Expected finish given preseason rank

From there, I thought it would be interesting to see where most of the preseason ranked teams finished. To do this, I first tried to examine the average finish for each rank from the preseason. However, I was unsure of how to handle teams that finished unranked.

I decided it would be simplest to just set all unranked teams to have a rank of 30. But then this would complicate the average ranking. They would be improperly weighted. An unranked team could be either have 4 losses or 10 losses, but would be weighted the same way in an average. Additionally, the "30" might be a bias against the better-ranked teams. The average is too sensitive to outliers to be too descriptive. If 9 teams finished at #4, and one finished unranked, would an average of 7 be that telling? Lastly, I wanted to just remove any possibility of looking at numbers greater than 25. If a rank's average finish

---

2[!] http://en.wikipedia.org/wiki/Spearman%27s_rank_correlation_coefficient

is at 28, what does that functionally mean? Not much, considering teams are only ranked 1-25 and "30" was an outside construction for the purposes of this paper.

Instead, I focused on median outcomes for each preseason ranking.

I found that the preseason rankings did a pretty good job of predicting penultimate outcomes. There were a few anomalies, such as #2 performing best, #13 performing so poorly, and #21 doing so well, but the general trend was clear. Higher ranked teams finish the season better.

Although I did not think average finish to be as useful, I still found the values just for the sake of comparison.

They pretty closely mirrored the results of the median finish.
Statistical Significance
The logical next step was to test whether the differences between these finishes was actually statistically significant. I ran a difference of means t-test to see if the average finishes for preseason #s 2-24 were statistically significant from #1. Unsurprisingly, at the 10% level of significance, all were significantly different except for #2.

Then, I ran the same test, except to see if the finishes were significantly different from that of the preseason #10. I did this because it may be particularly easy for preseason voters to determine the best two to three teams in the nation, but guessing the tenth best team from the beginning can be much more difficult. I found that the average for the preseason #10 was not significantly different from any of the other preseason rankings except for the preseason #1 and #2, which were significantly better, and the preseason #23, #24, and #25, which were significantly worse.
Bias
Now, knowing that #1 and #2 did significantly better than #2, I wanted to see how much of this I could attribute to bias. To test this, I found the average number of losses for each preseason rank. I used losses as opposed to wins because teams play more games and win more games now than they did ten years ago, but the best teams still usually finish with 1-2 losses.

On average, the top two preseason teams lose about three fewer games than the preseason ten. This justified the significant difference in finish. The top two teams finish better not because of preseason biases, but simply because they are better teams.

However, I did find all the teams in the middle to be interesting. The preseason #9, on average, only lost .5 games more per season than the preseason #3. But, as shown earlier, the preseason #9 tends to finish around #19 whereas the preseason #3 finishes around #3. I believe this to be some small evidence of a preseason bias, even if the average finishes are not statistically significantly different. Perhaps with more trials, this could be fleshed out better.

Median preseason rank for highest ranked team with X losses
I thought it might be interesting to look at the situation from the opposite perspective. I looked at the penultimate poll's highest ranked undefeated team, the highest ranked 1-loss team, highest-ranked 2-loss team, etc. I went back and examined their preseason rankings. My hypothesis was that teams ranked well in the preseason may rank worse in the penultimate poll due to underperforming, but compared to other teams with the same record, they would still have an advantage.

In general, I found this to be true. Most of the time, the highest ranked undefeated, 1-loss, 2-loss, and 3-loss teams were all ranked in the preseason top 10. This is probably my strongest case for a preseason poll bias.

Tracy Williams   December 10, 2014
Extended Abstract

Out of the 40 teams that were the highest ranked team with X losses between 2004 and 2013, only 6 were unranked in the preseason. Given that nearly 33% of teams in the penultimate poll were unranked in the preseason, there seems to be some favoritism for the teams that were ranked well in the preseason.

Bowl performance

The last thing I wanted to look at was bowl performance. How well does each poll predict the winners of the bowl games? Most bowls attempt to create a game between two evenly matched teams, so anything better than 50% would be interesting.

To my surprise, I found that the preseason poll did better in predicting winners of bowl games than the penultimate rankings. Despite the fact they contain very little information compared to the penultimate poll, on average, the preseason poll was correct in 63% of games in which it had a prediction. Meanwhile, the penultimate poll was only correct in 58% of games in which it had a prediction.

I ran a difference in means test to test the significance of this difference. The difference between the averages was not significant at the 5% or 10% level. However, just the fact that they are close was still surprising to me.

I conjecture that the preseason poll might have an advantage because it predicts different games than the penultimate poll. The set of teams in the preseason that make a bowl is different than the set of penultimately ranked teams that make a bowl game. As a result, the preseason poll might be predicting easier games. For example, in the upcoming Foster Farms Bowl, Maryland will play Stanford. Neither team is ranked, so the penultimate poll makes no prediction. However, Stanford was the #11 team in the preseason, so the preseason poll would predict Stanford to win. This may be an easier game to choose than a bowl between ranked teams because of Stanford's perceived superiority in talent based on its good preseason ranking.

Conclusion

On the whole, I did not find as much evidence as I had expected in support of the preseason polls biasing end of season results. Almost 1/3 of the final poll consists of teams that were unranked in the preseason. Only four preseason ranks (#1, #2, #5, and #6) average a top-10 finish. The preseason #1 and #2 finish significantly better because they finish with significantly fewer losses.

Still, it was interesting to see that teams with about the same number of losses could finish so differently in the rankings. Preseason teams 3-9, on average, finish with about the same record, but the preseason #3 tends to finish much higher in the rankings. Although the average finishes were not significantly different from each other, it would be interesting to further investigate this mechanism.

Additionally, the highest ranked teams with a certain number of losses tend to be ranked highly in the preseason. It seems as if as top teams falter, they still receive some boost over their peers with the same record.

Although my findings were not nearly conclusive enough to indisputably prove preseason poll bias, they still show that the question remains open.

Further Research

With more time, I think it would be interesting to test for several other factors in the polls.

A measure of strength of schedule could help control any conclusions about poll movements given the same record. It is difficult to devise an objective measure of strength of schedule, but it could help to better understand why some teams may rise more than others.

Additionally, there are AP polls for every week in the season. Some incorporation of these polls could add granularity to the analysis of poll movement.

Finally, I think it would be helpful to analyze for other biases. Does Notre Dame really get an unfair bump for being Notre Dame? Is SEC bias real? Conference and specific team biases could be tested against general preseason biases.