John Incantalupo

CSC-485 Capstone

May 10, 2024

Assessing the Value of NFL Punters

Introduction

Over the past decade, the sport of baseball has been overtaken by a new wave of analytics. The main statistic leading this revolution is Wins Above Replacement, or WAR. This all-encompassing metric seeks to summarize a baseball player's overall contributions to his team. The goal is to determine how many more wins the player's team won because they had that player as opposed to a "replacement-level" player, such as an average minor leaguer. This is based on how many runs a player theoretically creates or prevents by their hitting, pitching, defense, and baserunning. Ever since this metric was brought into the spotlight around ten years ago, statisticians have attempted to come up with similar metrics for evaluating players of other major North American sports, such as American football. However, due to the specialized nature of football, it can be hard to compare players from different positions.

Meanwhile, there is one position on the gridiron that stands out. The punt is one of the most unique plays in all of sports. It goes against the conventional wisdom of the sport, as you are seemingly giving up the opportunity to score by giving your opponent their own chance to put up points. By this logic, the value of a punter can be hard to quantify. So, I decided to create a WAR-like metric with the goal of measuring the performance of each NFL punter. I also wanted to see how my metric stacks up against other established measurements of punter performance. Since there are usually only 32 employed punters in the NFL at a given time, I was not able to create a “replacement-level” baseline for punters. Therefore, my metric will be comparing NFL punters to the average punter.

Data Overview

nflfastR is an R package that contains play-by-play data going back to the 1999 season. It features basic play information, such as the down and distance, the number of yards gained, and players involved, as well as more advanced metrics, such as expected points added and scoring probabilities. The most important variable for this project is kick distance, which gives the distance that the ball travels on a kickoff or punt before it is touched or goes out of bounds. Return yardage is not included in this metric, which I find useful. For the most part, I believe that a long punt return is something that is outside of the punter’s control and thus should not be incorporated into a punter’s overall value. Notably, the dataset was lacking a variable that gave the number of points scored on the drive. I was planning to find the average number of points scored per drive based on starting field position, so this variable would be important to have. As a workaround, I used the binary variables used for identifying scoring plays to add a new variable that gave the number of points scored on that particular play.

Methodology

To begin calculating each punter’s value, I wanted to find the value of starting field position. I grouped the dataset by each drive, as given by a unique numeric identifier for each drive in a game, in order to find the number of points scored on each drive. From there, I grouped each drive by starting field position and found the average number of points scored for each yard line. As predicted, a team is expected to score more points when they begin their drive closer to their opponent’s end zone, as shown in Figure 1. The relationship between a team’s starting field position and average points scored over the past 25 seasons produced a correlation

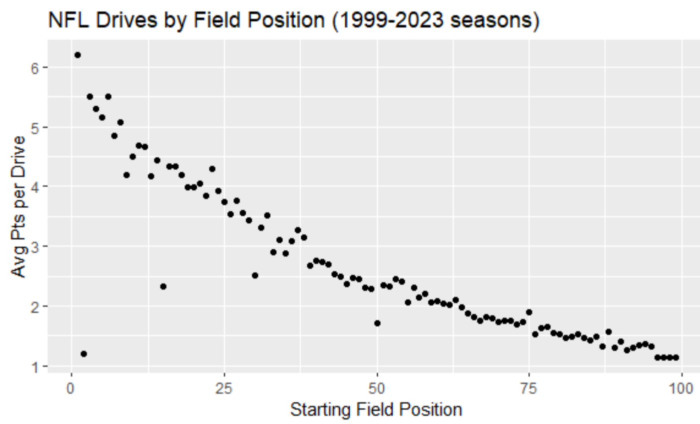

Figure 1

coefficient of about -0.9. However, there are a few outliers in the bottom left corner of Figure 1. This is unfortunately due to numerous errors where the drive number variable is updated one play too early. In many of these instances, that play was an extra point attempt, which occurred at the opponent's two yard line prior to the 2015 season and at the fifteen yard line during every season since. That extra point would be calculated into the following drive's total, which would often not result in a score. These "one-point" drives led to an unusually large sample of drives starting two yards away from the end zone with an unusually low number of expected points. Due to time constraints, I was unable to fix all of the errors and decided to continue on with the project. Since the number of drives with errors was a small percentage of the total number of drives, I felt that it would not drastically affect the results of this project.

Next, I calculated the average punt distance based on the current line of scrimmage during the punt play. This will give us our expected starting field position values, as I plan to compare those to the actual starting field position for each punt over the past 25 seasons. Figure 2 tells us that the average punt distance remains constant until close to midfield, where it steadily decreases as the line of scrimmage gets closer to the end zone. Even with the nonlinear pattern of the plot, it still produced a correlation coefficient of 0.89. Touchbacks were also taken into consideration, as I subtracted 20 yards from the distance of those punts in order to account for the starting field position being automatically placed at the 20-yard line following a touchback.

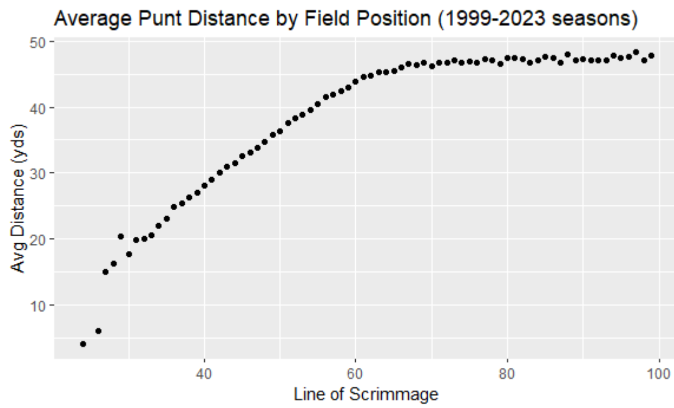

Figure 2

Now that we have the field position data from the past 25 seasons, we can measure each punt's value compared to the average punt from that yard line and attribute the value to the punter. First, I calculated the observed number of expected points for the opponent's next drive based on each punt's distance. Then, the opponent's expected starting field position was computed by using the average punt distance from the punt's field position, where expected points for the expected field position were also calculated. The difference between the two expected points values gives us points above average, or PAA. This hopes to give us a measurement of the number of opposing points that a punter theoretically "prevented" by kicking the ball farther than the average punter. By grouping PAA by punter, we can see which punter was the most valuable in putting their opponents in worse field position.

Once I had aggregated each punter's PAA and found both individual season and career totals, I wanted to compare my findings to a pre-existing, credible measurement. Pro Football Focus, or PFF, is an online resource known for their numeric grades of NFL players. PFF has a team of analysts that take each player involved in a given play and rate their impact on the play on a scale of -2 to +2. At the end of the season, these ratings are converted to a different scale ranging from 0 to 100, which represents the player's overall grade. This is one of the most robust attempts at creating a metric that can compare NFL players of different positions, including punters. We have access to PFF's punter grades going back to the 2013 season, as well as other punting metrics such as hangtime and average return yards.

Results

In baseball sabermetrics, it has become commonplace to incorporate a seasonal adjustment into various statistics in order to accommodate for different eras of the game. I wanted to do something similar with PAA but was unsure of how much of an effect it would have. I ended up calculating two different kinds of PAA: one where the opponent's expected points per drive were calculated from that particular season, and one where the expected points were taken from aggregate data from the past 25 NFL seasons.

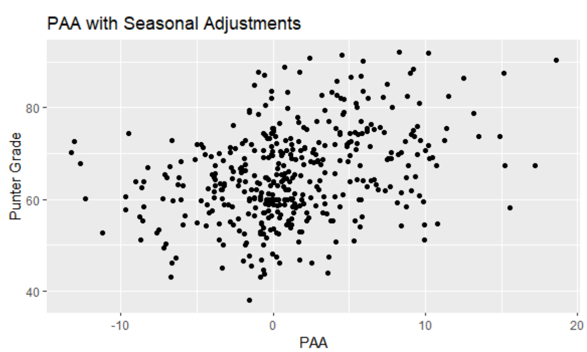

Figure 3

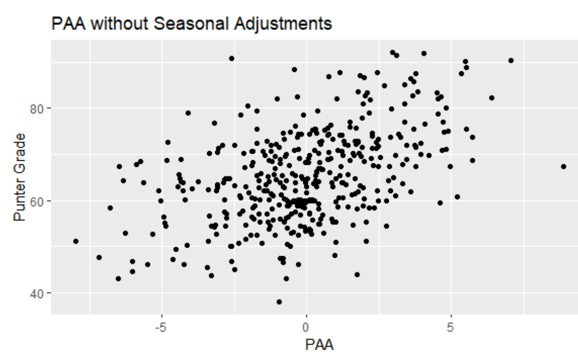

Figure 4

It immediately came to my attention that these two methods of calculating PAA would yield varying results when I found that the correlation coefficient between the two was 0.621. The variance of PAA was much higher when adjusting for the particular season, as PAA values ranged from -13.2 to 18.6. Meanwhile, PAA values without seasonal adjustments only ranged from -8 to 8.9. Figure 3 plots PAA with the seasonal adjustments against PFF's punter grades for every individual punting season since 2013. The correlation coefficient was a meager 0.357, indicating that there is not much correlation between the two metrics. The Spearman's rank correlation coefficient was also calculated for the two variables, and the coefficient essentially remained the same. Surprisingly, the PAA metric's correlation with PFF grades improves when we remove the seasonal adjustments, as shown in Figure 4. The coefficient rises to 0.483, which is a respectable number considering that the only metric going into PAA is punt distance. In theory, taking the season-to-season variance into consideration should control for the league's offensive environment. However, the smaller sample sizes for each season's field position data negate this variance, leading to a worse correlation. Looking at Figure 5, which gives the average number of points per drive grouped by starting field position for the 2023 season alone, the data points are more jittery than the ones we saw back in Figure 1.

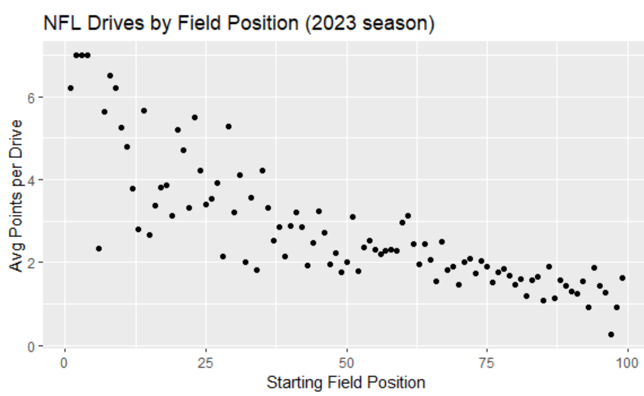

Figure 5

Although the overall correlation between PAA and PFF grades was moderately weak, I did want to see if there were any specific seasons that yielded a strong correlation, particularly using the PAA metric without seasonal adjustments. The year 2016 produced the highest correlation coefficient, which was 0.645. After incorporating PAA with seasonal adjustments into these calculations, I found that the PFF grades had a better correlation with the PAA metric that did not include seasonal adjustments in each of the 11 seasons with data, as shown in Figure 6.

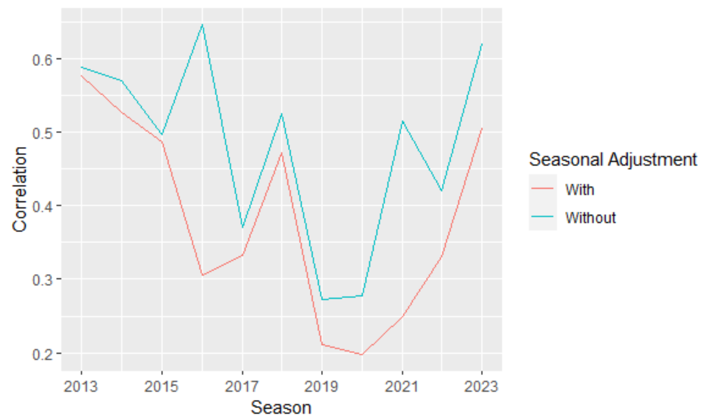

Figure 6

<u>Conclusion</u>

Overall, the moderate correlation between PAA and PFF's punter grades does not make PAA alone a suitable measurement of an NFL punter's value. Although the errors with the drive number variable definitely did not help, I do think that the main cause of the relatively low correlation is PAA's sole reliance on punt distance. Using R's corrplot package, I created a correlation matrix (see Figure 7) containing many of the variables found in the PFF dataset in an attempt to determine what metrics PFF values the most when grading punters. The metric that stands out the most is hangtime, as it has a correlation coefficient of 0.619 with PFF grades. This

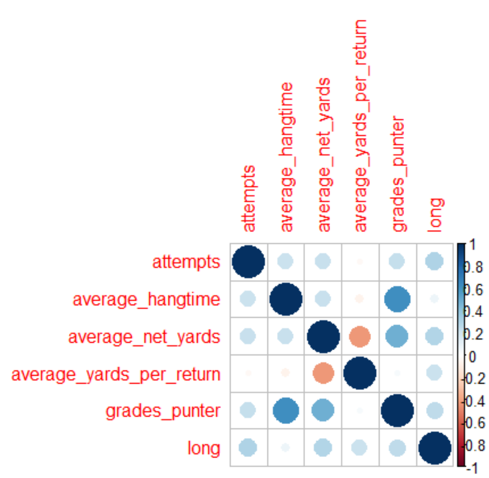

Figure 7

is a metric that I would have wanted to include in my calculation of PAA if it was available in the play-by-play dataset. The ability to keep the ball in the air for as long as possible to give the coverage team more time to swarm the returner is a valuable skill for punters to have. If I were to take this project further, finding an alternative dataset that includes the hangtime for each punt is a priority.

The issue of positional specialization will always be present when it comes to creating value-based metrics in the sport of football. Pro Football Focus has done an invaluable job at creating a system that can be used to compare players who play vastly different positions, even if it requires hundreds of hours of film study from analysts. The NFL has seen its own analytics revolution over the past few years, with AWS's Next Gen Stats being brought to the forefront of the league. However, PFF's player grades predate Next Gen Stats by almost a decade. With this influx of new analytics, the way that players, fans, and coaches view the game has drastically changed. Even still, statisticians and analysts continue to scour data in search of even better ways of measuring on-field performance. With that being said, I firmly believe that one day there will be a WAR-like metric for measuring the value of any NFL player, and with that, we can finally see the true value of the punter.