The Effectiveness of Professional Punters

Arthur N. Tanyel

A Senior Thesis submitted in partial fulfillment
of the requirements for graduation
in the Honors Program
Liberty University
Spring 2021

Acceptance of Senior Honors Thesis

This Senior Honors Thesis is accepted in partial fulfillment of the requirements for graduation from the Honors Program of Liberty University.

____
David E. Schweitzer, Ph.D.
Thesis Chair

____
Timothy Van Voorhis, Ph.D.
Committee Member

____
Emily C. Knowles, D.B.A.
Assistant Honors Director

____
Date

Abstract

Sports analytics have become a major part of how many sports fans enjoy the games they love. The trend has touched sports from football to cricket. One aspect of football that has been less discussed is punting. The current standard metrics, gross yardage and net yardage, give an idea of how a punter performed in a given season, but they also may be skewed by a number of factors, such as the skill of the offense or the punt coverage team. In this paper, we will look at some previous attempts to measure punting prowess, and then further develop a metric to reward punters for their skill while accounting for uncontrollable factors such as long returns and penalty yardage.

The Effectiveness of Professional Punters

The Sports Analytics Revolution

Sporting events and competitions have been a part of human life for so long that their origins have been shrouded by the passage of time (Kyle, 2014). Despite this, statistics in sports are a relatively new phenomenon. Through the nineteenth century, sports grew more structured with the rise of school sporting events first and then highly organized amateur competitions (Holt, 1989) such as the Football Association in England and the National Association of Base Ball Players in the United States. In these more complicated team sports, players, managers, and fans looked for ways to better measure and understand the skill of competitors. It is simple enough to see how effective a wrestler or runner is. It was quite obvious, for example, that Aquilo, a Roman chariot horse, was successful because he, “won 130 races for the Red circus faction” (Golden, 2004, p. 14). He won, and so he was good. However, in team sports such as baseball, football, rugby, or basketball, it can be more difficult to determine the value of individual players. As a result, certain statistics began to be tracked and others were invented. Hits in baseball were recorded. To give a better comparison of success between players, the batting average was invented, dividing the number of hits by the number of opportunities to hit. After organized football began in the 1920s, players were given credit for the number of tackles they made or the number of yards they gained in various manners.

For a long time, these statistics remained this simple. While the National Football League (NFL) made sacks an official stat in 1982 (Weinreb, 2010) and Major League Baseball (MLB) began tracking on base percentage in 1984 (On-base percentage, n.d.), these additions were just minor adjustments to existing formulae. However, it would take outside individuals to begin developing more advanced sports statistics. One such effort, whether in an attempt to better quantify candidates for the Hall of Fame or the search for better team building tools, was the Society of American Baseball Research (SABR), founded in 1971 (Costa et al., 2008). From there, the world of sports analytics expanded in every which way, enveloping every major modern sport from football to cricket.

Sports Analytics in Football

One of the challenges that football presents is that it is a significantly more complicated game than baseball. While batting, fielding, or pitching achievements are largely attributable to an individual, each play on a football field heavily involves many players (Pasteur et al., 2018). At the very least, on any given play the failure or success of any of the five offensive lineman, three or four defensive lineman, and the quarterback are all intrinsically linked with the failure or success of the play. Furthermore, the relative success of a play is related to the situation it occurs in (Chait, 2004). Clearly, a first down late in the fourth quarter of a close game is much more important than a similar play in the first quarter.

Despite the challenges, there have been many efforts to quantify football performance. One of the earliest popular examples was developed by Football Outsiders, an advanced metrics company, in 2003. They introduced Defense-Adjusted Value Over Average (DVOA), their method of evaluating teams, units, or players (Methods, n.d.). Essentially, DVOA takes the outcome of a play and compares it to the league average outcome in similar situations. This was followed by similar ideas such as Expected Points Added (EPA) and Win Probability Added (WPA), created by Advanced Football Analytics. These statistics also measure the value of particular plays. We will discuss EPA at a later point, but WPA takes the difference between a team's probability to win before and after a given play (Burke, 2014a). If, for example, a given pass increases a team's win probability from 0.44 to 0.47, that play is assigned 0.03 WPA. It should be noted that neither of these statistics attempt to assign credit to individual players. It is the outcome of the play as a whole that is assigned WPA.

Other efforts have been made to evaluate individual players. One such attempt by Advanced Football Analytics is the Tackle Factor (TF). TF is the percentage of a team's total tackles that an individual defender is credited for, adjusted for the number of games played (Burke, 2014b). This is a fairly simple and intuitive statistic, but it has its limitations. One such limitation is that TF "tells us something very different about defensive backs than for lineman and linebackers" (Burke, 2014b, para. 8). If a cornerback has a significantly large percentage of a team's tackles, it may indicate he is not very effective in pass coverage. For a safety, a high TF may mean the defense in front of him is letting a higher proportion of runners get to the second level. These issues highlight just how difficult it can be to assign single numbers to a football player's output due to the complexity of the game.

The one possible exception to this rule appears to be field goal kicking. Osborne and Levine (2017) employed several weighting schemes for a shrinkage model used to assign value to individual kickers. They found that in general, when the stadium was taken into consideration, their results agreed with a kicker's field goal percentage. This is likely due to the fact that relatively few players influence a field goal attempt. The snapper and holder must both perform their jobs well, but at the NFL level, it is very rare that any of the offensive blockers allow the defense to influence the kick. That being said, the 21 non-kicking players on the field still may have an impact, even if it is very rarely.

Punting Metrics

While there have been metrics developed for nearly every aspect of football, one of the more neglected areas has been punting. This is possibly because it is one of the least glamorous activities, perhaps only behind long snapping. The two main methods of measurement, gross punting yards and net punting yards, are basic and relatively intuitive. Gross yards are simply a measure of how far the ball traveled past the line of scrimmage. Net punting yards are only slightly more involved; simply subtracting from gross punting yards any return yardage the receiving team gains. These naturally give rise to gross yards per punt and net yards per punt. Both of these statistics are quite flawed as they "are significantly impacted by field position, and also fail to take into account hang time and the directional ability of a punter" (Zodda, 2016a, para. 1). Additionally, there are several other factors out of a punter's control that may affect his statistics.

Two additional traditional metrics exist. In box scores, punters are given credit for the total number of their punts that land inside the 20-yard line. The other measure, hangtime, is generally not officially reported, but is often referenced by broadcasters (Pasteur, et al., 2018). A longer hangtime indicates that the punted ball is in the air longer, and so affords the punter's teammates more time to get into position to make a tackle. Much like gross and net yardage, these numbers do provide usable information, but they must be treated with caution as several non-punter factors can influence them.

Punting Factors

All four of the above statistics may be affected by the quality of the team around the punter, either for positive or negative outcomes. There are two main aspects of the team around the punter that may affect punters. Neither of these are accounted for by traditional metrics.

Relative Strength of the Offense

A relatively good or bad offense leads to different punting scenarios for punters, which results in varied gross and net punt yardage. A punter with a weak offense is more likely to find himself punting from deep in his own territory. This allows for large gross punting numbers, as the punter can simply focus on kicking the ball as far as he can. A punter with a stronger offense, however, is more often faced with the task of pinning his opponents deep in their own territory. The punter in this case has much less field to work with. He also must try to avoid kicking the ball into the endzone, which results in the ball being placed at the 20-yard line. This, while not a tragedy, is not the most ideal outcome.

Additionally, a punter with less space to work with can focus more strength on getting a ball high into the air, while if he is backed up on his own side of the field, that strength must be used more horizontally. In this way an offense can affect average hangtimes. Thus, two punters with varyingly successful offensive teams will likely have different statistics, even if they have the same level of skill.

Skill of the Punt Coverage Team

In a similar way, a better punt coverage team will increase a punter's net yardage in two ways. To begin with, a better coverage team will tackle the returner sooner, limiting how many yards are taken off from each punt. Additionally, they may be more skillful at downing the football before it bounces into the endzone for a touchback. This clearly can have an effect on the number of punts inside the 20 a punter has, but skills in both of these areas also could lead to an increase in a punter's net yardage in a way that does not account for the punter's skill, while deficiencies could unfairly punish the punter. Zodda (2016a) gave an example showing that there are instances where net punting has almost nothing to do with how the punter performed.

While this is anecdotal evidence, the play in question does show that the punter can do his job quite well while not receiving any credit from traditional statistics.

Playing Conditions

The weather in which the game is played can have a significant effect on punts. Colder or warmer weather will affect both the pressure of the ball and how well it can fly through the air. Rain, snow, or mud can slow down the ball and the punter. Perhaps most significantly, wind can drastically change how similarly kicked balls fly, both for good and for ill. Combine these direct consequences with any effects weather may have on the offense or punt coverage team, and weather may be the most significant factor when it comes to punting.

Additionally, the altitude at which the game is played may have an effect on how far balls fly. Naturally, a ball punted in a lower pressure environment will travel farther than a ball punted in a higher pressure one when kicked with the same force. This may seem insignificant, and it may well be for most NFL stadiums, but at the extremes, there may be a difference. The New Orleans Saints home stadium is nearly a mile lower than that of the Denver Broncos. This may result in a perceivable difference between equally skilled punters.

Value as Punts Near the Goal Line

Not every yard on a football field is of equal value. Clearly, gaining the last five yards before the endzone is of more value than gaining five yards in the middle of the field. Similarly, a punt that travels only 40 yards, but lands on the opponent’s five-yard line is more valuable than a punt that travels 40 yards and lands on the opponent’s 15-yard line. Properly valuing these yards is one of the main difficulties of valuing a punt. This is where the idea of target lines begins to have merit. If that target line is created by some sort of league average, we are able to get some idea how much value a specific punt provides. That being said, a performance that is

5% better than average near the goal line may still be of greater significance than a similar performance in the open field. However, despite the difficulties, there are useful and innovative ways to account for this.

Analytic Attempts to Quantify Punting

With these factors in mind, several statistics have been developed to account for them. The authors of these metrics range from academics to fans with an intense interest. The following are some interesting efforts.

Expected Points Added

We must begin our discussion with Expected Points Added (EPA), which is traditionally an incredibly useful metric, but also a terrible way to measure a punter's skill. Romer developed a method to calculate how many points on average certain decisions would add to a team's score (2006). He did this in an effort to quantify when teams should attempt to convert fourth downs. As a result, he had to quantify the alternative, which, in one case, means punting the ball away.

After designing a function to value expected points added from virtually every scenario, he netted the opponent's expected points against the kicking team's expected points if they had converted, or failed to convert, the first down. This leads to a measure of how much the decision to punt affected the game. However, giving the ball away, no matter how prudent the decision is, will almost always lead to negative EPA value, particularly on fourth and one. The odds of converting a first down in such a scenario are so great, while the effects of giving the ball away so negative, that no matter what the punter does, EPA does not favor him. Furthermore, Romer's methods account for the time of the game, the current score, and other factors that are out of a punter's control (EPA/punt, 2020). This is largely because EPA was developed not as a way of measuring punting ability, but as a way to help guide a coach's decision-making process.

EPA does not value specific punts based on their distance and landing points. It compares the value of having the ball at a specific location versus the opponent having the ball at another location, but it does not value the punter's work based on where he kicked the ball from. Neither does it award the punter independently of any other forces that are outside of the punter's control. However, this measurement has been used elsewhere in conjunction with other means to help quantify a punter's success.

A Flight Based Metric

Pasteur et al. (2018) used Romer's work in an effort to quantify the value of punts as they near the goal line. They charted the results from all punts in the 2013 season visually, using the NFL's online video service to track landing points, ending points after bounces, hang time and other factors. They also collected data about the weather for each game and added any penalty data onto the result of the play that was not listed in their play-by-play sources. Using this data, they model the effects that these elements have on punts. For example, they found that an average punt in the NFL in 2013 included 0.86 penalty yards against the return team. Doing similar calculations, the authors endeavored to eliminate confounding factors such as the aforementioned penalty yards, return yards, distance the ball bounced and rolled after its flight, and any weather effects.

Once these confounding factors are added or subtracted from a punt, the landing spot of the kick is compared to the league average landing spot of balls kicked from the same yard line. Specifically, the Expected Points of an opponent's drive from the expected landing spot is subtracted from the Expected Points of the adjusted landing spot. These Expected Points were both aggregated over the whole season and used to calculate Expected Points per punt.

Tymins' Expected Points

Tymins (2014) also developed a method to use Romer's EPA to measure punting success. To begin, he first found the league average yard line for the next play from scrimmage for a punt from each yard line. This gives punters a target against which their individual punts may be compared to. Already seen in Pasteur et al.'s (2018) method, this is a common idea that takes several forms and will be returned to later in many other statistics. As previously noted it does not suffice to compare the target line to the actual outcome, as each yard of a football field is of differing value. Much like the previous study, Tymins (2014) nets the expected points of a drive that starts at the target line against the expected points based on where the actual drive began. Again, these expected points can be accumulated over a season or examined on a per punt basis.

The Punt Runts

The Punt Runts are a group of Yale University students who have taken to developing punting metrics in their spare time (Puntalytics, n.d.). They use nflfastR, a package of functions used to analyze NFL play-by-play data (nflfastR, n. d.) to develop insightful looks into punting skill. We examine two of their efforts

SHARP

The first attempt by the group was scrimmage help/hurt adjusted real punting (SHARP), a metric that measures a kick by comparing it to the league average from each yard line ("Gross yards," 2019). After smoothing the average to eliminate noise, SHARP is determined by dividing the gross yardage a punter earns on a kick by the league average gross yardage from the original line of scrimmage. Thus, if the league average from a yard line is 50 yards and a kick travels 40 yards, the punter is awarded a SHARP score of 80%. Over the course of a season, a punter's average SHARP is the main method of evaluation under this system.

The main limitation of SHARP comes from the fact that it is inherently linked to gross yardage, and therefore does not account for punt returns (SHARP, 2019). To challenge this issue, net yardage may be used in much the same way to create SHARPnet. The combination of these two statistics can give a decent idea about how punters rank relative to each other.

EPA/Punt

Following SHARP, the group developed a statistic based on Romer's EPA called era-adjusted EPA above expected per punt. EPA is the standard statistic for many metrics in other areas of football, but, as detailed above, it is not inherently useful for punting (EPA/punt, 2020). They begin by averaging the EPA for a fourth down of any distance from a specific yard line. For example, they found a fourth down on one's own twenty-yard line, aggregated for all distances to gain is worth -2 expected points, on average. Armed with these averages, they then got a better comparison between the team's situation before and after the kick. Thus, if the resulting yard line gives the kicking team a situation where they have -1.3 expected points, that number is subtracted from the original situation, yielding EPA for the punter.

This number was taken and compared to the league average of EPA from the yard line in question. Thus, the punter earns EPA above expectations. This helps eliminate the negative impact of a coach's poor decision that traditional EPA would assign unfairly to the punter. Rather than comparing the objective value the punter provided the team (which in some cases can be negative), the statistic compares how the punter performed relative to other punters put in the same situation. This leads to a metric that uses the team-centric EPA in a way to measure punters specifically.

Target Distance Punted

Target distance punted (TDP) measures a punter's skill by comparing each punt to the league average in whatever situation he is punting from (Zodda, 2016b), much in the same way SHARP does. Developed by Chuck Zodda, a writer for Inside the Pylon and former kicker at Dartmouth College ("Author", n. d.), TDP first separates each punt into two categories: pin-deep situations, which occur when punting from on or beyond the kicking team's 41-yardline, and open-field situations, which occur behind the kicking team's 41-yardline (Zodda, 2016c). In open field situations, the calculation of TDP is much the same as SHARP. So, a punt expected to travel 50 yards that travels 55 receives 110% TDP, just as it would under SHARP.

In the pin-deep case, the target is no longer set at the league average result. Instead, the distance of the line of scrimmage from the punter's goal line is subtracted from 90. So, the target distance when punting from the opponent's 45-yardline would be 35 yards. Again, the gross distance, less any yardage lost on touchbacks, is compared with the target. So, a punt from the 45-yardline that flies to the 10 would have a 100% TDP while punting the ball only to the 20 would result in approximately 71% TDP. The authors do not discuss using net yardage in the same way, but considering the example of SHARP, this may be a reasonable extension.

The Cowardice Machine

Bois (2019) documented his quest to find the most cowardly punt in recent NFL history, per his Cowardice Score. Trawling through all regular season punts from 2000 to 2019, he judges each punt based on where it occurred on the field, what the score was at the time, how much time was left in the game, and other situational factors. This score, like Romer's work, is more useful for deciding whether a decision was a good one or not rather than assigning success or failure to a given punter.

A New Statistic to Measure Punts

In light of the existing metrics, we believe there is more room for exploration. Many of the metrics above are useful, but they have some limitations. Romer's (2006) EPA, while a very useful tool, is not enough on its own to measure the success of punters. It must be developed more thoroughly to be useful.

We found the method developed by Pasteur et al. (2018) to be easily the most compelling use of EPA. They also covered the many different confounding factors admirably and effectively. The only limitation to their study is that it is quite time consuming and not easily replicable by novices or even industry insiders. Computer code can be used to aggregate play-by-play data, but the information they used involving weather and especially hang time was much more involved. If one wanted to conduct a study by their methods over the previous five years, they would need to take 3 measurements for each of the approximately 10,000 punts launched over that time span. With motivation and money, it is certainly a doable task, and the task would be made simpler if the NFL collected the data themselves as it was created, but we find it difficult to imagine such an interest, either intellectual or monetary, exists for punting.

Tymins' (2014) effort has the advantage of being more accessible than the previous attempt. It retains the key use of EPA as the measure of comparison between target and reality, while eliminating the many hours required to find weather, hang time, and bounce data. This naturally leads to a weaker model, but depending on the audience, that may be a sacrifice worth making. However, there is not much of an effort given to measure how well a punter did on a specific punt. A punt that is returned for a touchdown, or even returned unusually far, may not be the punter's fault, but would still yield a poor evaluation from this method. Now this cannot be completely accounted for, but there can be improvement. Furthermore, Tymins (2014) was not clear in how he calculated EPA for each yard line. By its nature, EPA takes into consideration the score and time left in the game when assigning value. Thus, it would not make sense to compare a punt in the fourth quarter to one in the first. It seems likely that Tymins (2014) accounted for this, but it is not commented on in his work.

EPA above expected per punt is a useful, but not as strong an attempt as Tymins' (2014) work or Pasteur et al.'s (2018) flight-based metric. It is another useful way to assign credit to a punter using EPA, but it ignores many of the factors previously discussed that are out of the punter's control. Specifically, using the next line of scrimmage fails to account for how any returns or penalties may have affected the outcome of the punt. While a punter may be complicit in allowing a large return, it is certainly never his fault alone. Thus, while useful, it is not entirely effective in assigning value to just a punter. It is perhaps more suited to assigning value to a punting unit.

The other effort from the Punt Runts (2020) suffers from a similar problem. Although not linked to EPA, SHARP and SHARPnet both use the next line of scrimmage to assign value without taking into account any other factors. So, while SHARP and SHARPnet are useful and intuitive metrics, they can be improved upon.

TDP, like the metrics developed by the Punt Runts (2020) has a similar limitation. They once again compare an outcome to a specific target without considering returns or penalties. Furthermore, their target calculation in pin-deep situations seems a bit arbitrary. It certainly works since every punt in this category is treated in the same way, but why subtract from 90? Why not 85 or 80? Once again, this is not to say the statistic is useless, but only to point out a potential short coming.

A common theme appears to be present with the statistics. While they use some sort of league average to determine a target for a punter, they do not account for any individuals the punter may be dealing with. The returner may be particularly skilled, or a defender may make an error that is quite out of the ordinary. Furthermore, penalties can have a significant impact on where the receiving team starts with the ball. Pasteur et al. (2018) controlled for these, but the rest of their work is a bit inaccessible. Thus, we believe there is room for a statistic that attempts to control for these highly variable factors. However, the above metrics are not completely without merit. Therefore, we propose a statistic that produces a target line while also accounting for uncontrollable factors on individual punts.

Method

Using official gamebooks from NFL.com, we charted the outcomes of all 2,147 punts from the 2019 regular season, excluding blocks and tips. This included noting kicked distance, any return or penalty yardage, and changes due to muffs and fumbles. Using this data, we developed two methods to model the expected outcome of each punt, which was then compared with the actual outcome to assign value.

The General Method

Our first goal is to establish a target line for the line of scrimmage on the play subsequent to the punt. To do this, we simply found the average new line of scrimmage league-wide after a punt from each individual yard line. This is much the same approach as TDP, SHARP, or Tymins' (2014) method. However, due to confounding factors such as penalties and returns, it is not sufficient to compare the actual outcome to this target. Therefore, we attached adjustments to the actual line of scrimmage by adding or subtracting the league average for return yards, penalty yards, and changes due to muffs or fumbles. Table 1 contains a description of our adjustments, where positive yardage favors the kicking team.

Table 1

Summary of Adjustments

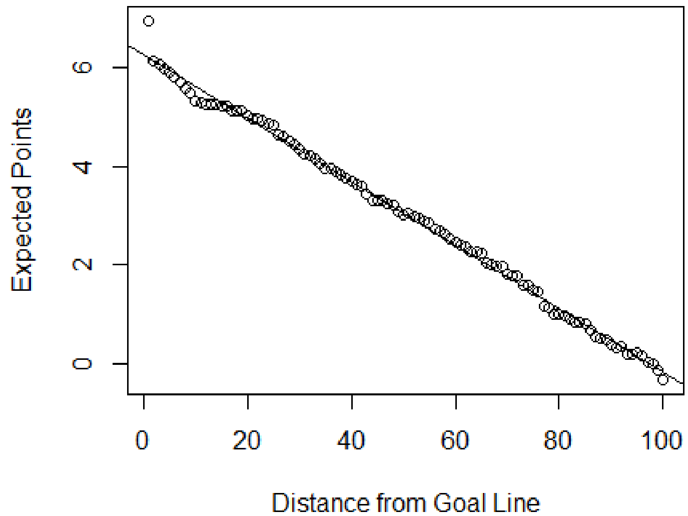
Variables	Adjustments
Return yardage	3.50
Penalty yardage	-0.77
Muffs and fumbles	-0.01

Functionally, this leads to a 2.72-yard adjustment to the actual new line of scrimmage. These adjustments help credit the punter for his work alone, while minimizing some of the effect of his teammates and opponents on each specific punt.

It is once again not enough to compare the adjusted line of scrimmage to the target line. This is because, as previously noted, not every yard is equally valuable. Much like Tymins' (2014) work, we compare the Expected Points of the opponent's drive based on the two starting locations. Specifically, we used the expected points for a drive from each yard line with no time run off of the clock in the first quarter. This eliminates the factors such as how much time is left and what the score is that are used in usual EPA calculation. Thus, a punt from the 25-yard line is expected to give the opponent the ball on their own 27-yard line with 1.59 Expected Points for that drive. If a punt leads to a new line of scrimmage on the opponent's 20, the opponent has 1 Expected Point for their drive. Thus, we would award the punter 0.59 Expected Points Above Average (EPAA). Figure 1 displays the number of points expected based on a first and ten at a given yard line.

Figure 1

Expected Points v. Distance from the Goal line



Note. The line of best fit is given by $y = 6.27 - .06x$

The Specific Method

Our second method is much like the first, but it tries to account for the fact that a blanket adjustment may not be prudent. For example, a punter is much more likely to have a kick returned if he is punting from his own 1-yard line than from the opponent's 40-yard line. In the latter case, the most likely outcomes are touchbacks, the ball being downed by a teammate, or a fair catch. Thus, it seems useful to calculate adjustments to be made for each yard line. In every other way, this approach is the same, but the change in adjustments makes a noticeable difference. Table 2 displays the top ten accrued EPAA over the 2019 season by starting punters through both the general and specific method.

Table 2

Top ten punters of 2019 by EPAA

Rank	Punter	Specific Season EPAA	EPAA/punt	Punter	General Season EPAA	EPAA/punt
1	Way, T.	13.97	0.176	Way, T.	14.05	0.178
2	Kern, B.	9.93	0.127	Kern, B.	11.23	0.144
3	Dixon, R.	8.71	0.126	Anger, B.	8.99	0.199
4	Huber, K.	7.91	0.105	Dixon, R.	8.83	0.128
3	Anger, B.	6.49	0.144	Huber, K.	7.1	0.095
6	Hekker, J.	5.95	0.090	Hekker, J.	6.92	0.105
7	Cooke, L.	5.19	0.069	Cooke, L.	6.88	0.092
8	Johnston, C.	3.81	0.053	Johnston, C.	5.27	0.073
9	Martin, S.	3.39	0.045	Morstead, T.	4.61	0.077
10	Morstead, T.	3.21	0.054	Colquitt, B.	4.59	0.075

Discussion

We believe this method of evaluating punters is intuitive, insightful, and relatively simple. It helps control for factors out of a punter's control, while also being accessible to anyone with a background in R using nflfastR. Even this is not necessary, as we did our calculations using Excel, but for anyone looking to investigate further, we highly recommend R, as it saves immense amounts of time and energy. While not being as in depth and Pasteur et al.'s (2018) methods, this relative ease of access is important for non-academics to use, understand, and perhaps modify the approach.

With that in mind, it is useful to examine the context that surrounds these statistics. If, by our calculations, Tress Way contributed about fourteen points to his team over the season, how valuable is that? An average NFL team scored 365 points over the course of the 2019 season (2019 NFL Records, 2020). Thus, Way, on an average team, could be credited with providing just under four percent of his team's production. For context, Peyton Manning produced approximately 43% of his team's production in 2013 (Tymins, 2014). Considering how many more snaps a quarterback is on the field for, this is not insignificant production from a punter. Now, this is not an exact measurement, as a punter's primary value is preventing the opponent from scoring rather than aiding the offense, but it gives an idea of the scale these metrics are dealing with. A fourteen-point swing over the course of the season is certainly of value. Additionally, the nearly thirty-point difference between the best and worst punters by our metrics is, based solely on how close most NFL games are, likely to affect the outcome of multiple games per season. Examining how NFL teams value this production is an interesting study.

The most obvious way to approach how a team values a player is how much they pay them. This is a bit of an oversimplification, but if a player is paid more, they must be valued more. The most glaring exceptions to this rule are players still on their rookie contract. The rules for rookie contracts are dictated by the NFL's collective bargaining agreement with the NFL Players Association, and they lead to a slotted amount the rookie can agree to, based on the position he was drafted in. This contract lasts four or five years and can lead to a player who is quite important to his team not being compensated as such. We will ignore players on rookie contracts in the discussion for this reason. It should be noted that only 18 punters in 2019 were not playing on rookie contracts (2019 punter salaries, 2020). If, as some people suppose, punters are largely replaceable, this is to be expected, as teams have to pay very little for rookie talent.

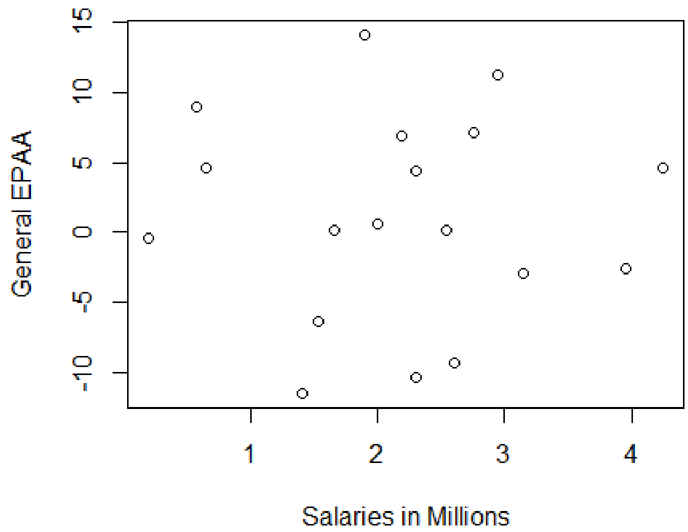
However, this may also simply reflect how short NFL careers are, meaning many players never get to their second contract.

Using salaries provided by Over the Cap (2020), a business specializing in providing NFL salary cap data, we investigated the earnings of veteran punters. The most surprising find on a cursory inspection is that Way is 12th of the 18 veteran punters, making less than fifty percent of the highest paying contract, despite being the best punter as measured by both general and specific EPAA as well as EPA/punt. The highest earner, Thomas Morstead of the New Orleans Saints, was not even in the top ten in either of our calculations. Another striking observation is that the next two highest earners, Sam Koch and Dustin Colquitt, both had negative EPAA through both the general and specific method.

In general, there seems to be little to no correlation between NFL spending, and a punter's performance. This may be due to NFL teams valuing things not measured by our statistics, such as experience or leadership, but it may suggest that there is an inefficiency in the punting market. Teams may want to consider paying some players more than what they are currently making, while others may realize they have awarded a contract to a player that does not match his on-the-field performance. Figure 2 gives a scatter plot of General EPA against salaries. It can be quickly seen that there is very little order in the plot.

Figure 2

General EPAA v. Salaries



Similarities to Other Work

Of note is the similarity of our rankings to those generated by EPA/punt. Both studies used the 2019 season as their data set, and in general, both metrics agree on above average and below average punters. While there is naturally some variation, both metrics put Way and Brett Kern at the top of the list, while Corey Bojorquez and Chris Jones performed at or near the bottom of the league. This similarity suggests there is merit to both methods and opens an opportunity to consider how the NFL values these punters.

Furthermore, while unable to directly compare players' performance, there is a similarity between our methods and those of Tymins (2014), as could be expected given their relationship. Specifically, the best and worst punters in Tymins' (2014) study produced similar numbers to ours. This, combined with the similarity of our results to EPA/punt, makes it seem very likely that punters are responsible for somewhere between -15 and 15 points for their team over a season.

Limitations

The first and most obvious issue is with the specific method. In many cases, the sample sizes from each yard line likely are not big enough to make a meaningful model. Most egregiously, there was only one punt the entire 2019 season from the opponent's 30-yard line. In fact, no spots before the kicking team's 10 or beyond the opponent's 40-yard line had more than 19 punts to draw averages from. Between the 20 and 50 yard lines, this is not as big a problem, with between 40 and 70 punts from each yard line, but as you get closer to either endzone, the sample size drops off.

A sample of multiple years would help fix this problem, but likely only to a certain extent. In the modern NFL, it seems very unlikely that researchers would get more than ten punts from the opponent's 30 over even a ten-year span. Furthermore, opening up such methods to longer time periods can allow new confounding factors like rule changes and new playing styles or techniques to take effect. As an example, it is not wise to compare a modern kicker's extra point percentage to a kicker from 20 years ago. The rules have changed to (in theory) make an extra point more difficult. While such an effect may be less pronounced, punting has still changed in significant ways in the last 20 years (EPA/punt, 2020). Therefore, this may be a necessary limitation of this method.

That being said, the method is not without some merit. The average difference between punter's seasonal general and specific Expected Points is only 0.88 EPA, and the general method does not suffer from the sample size issue. Furthermore, the overall rankings did not change significantly. The largest swing was experienced by Mitch Wishnowsky, who lost 4.2 EPA from general to specific and dropped from 15th in the league to 27th. Wishnowsky was an exception however, as he and Kern were the only players to experience a shift of more than 2 EPA. In addition to the method being imperfect, there were factors that affect a punt's outcome that we did not consider.

Areas for Further Investigation

While we believe our metric is generally useful, there are some specific aspects of punting that we were unable to measure. Some of these things might have fit well in our model, such as how near the sideline a punt is fielded, but due to our data collecting restraints, we were unable to include it. Others, however, such as catchability likely need their own investigation.

Punts Near the Sideline

One skill that every metric discussed does not consider is directional punting. To begin with, a punt that goes out of bounds cannot be returned. It also cannot bounce or roll in an advantageous direction, but given the shape of a football, predicting the way it will bounce is difficult, if at all possible. However, it also seems reasonable to assume that a punt returner has less space to work with when fielding a kick near the sidelines. Thus, it is possible that a punter may mitigate some return yardage based on how near the sideline he places his punt.

It seems likely that this factor has been ignored by so many metrics because the data needed is not readily available in play-by-play descriptions of games, either in information provided by the NFL or by nflfastR. A study of this sort would have to take a similar shape to that of Pasteur et al.'s (2018) work. While they used NFL film to calculate hangtime and other data, another party could use the same service to chart where each punt landed with respect to the sideline. While this would take a large amount of time and effort, it is an unexplored factor that may lead to a new, better understanding of punting effectiveness.

Catchability

New England Patriots' head coach Bill Belichick has a well-documented affinity for left-footed punters. In his time as head coach, the Patriots have employed almost exclusively left-footed punters (Vrentas, 2018). With the natural scarcity of such punters, this cannot be a coincidence. One common explanation is that a ball kicked with the left foot spins in the opposite direction a right-footed kick would. Thus, a returner who is used to reading right-footed spin may have more difficulty tracking a lefty's punt, leading to an increase of muffs (Vrentas, 2018). This may or may not be true, but it does suggest a question: can punters control how catchable their punts are?

Johnny Hekker, the punter for the Los Angeles Rams, appears to experiment with such ideas. Punters have long known about what they call a "banana punt" (Bien, 2018, para. 8). Essentially, a ball can be dropped in a very specific way such that after it is kicked, ball curves as it travels down field. This kind of kick is difficult enough that it is almost never used in a game situation (Bien, 2018). This also suggests that a returner would have much more difficulty tracking and fielding this sort of punt, as they must rarely see such kicks.

Finally, one more tactic a punter can employ is the knuckle punt. Much like a knuckleball thrown by a baseball pitcher, a knuckle punt is a punt kicked in a way that sacrifices velocity but can cause the ball to "change direction probably three or four times within the last ten [to] fifteen yards [of flight]" (Petchesky, 2013, para. 3). Clearly, such action would make fielding this punt a much more daunting prospect. Thus, it may be that field position is not all that should be investigated when considering a punter's worth. When examining one of Hekker's better banana punts, just considering the box score would tell you that it is a solid, above average punt. However, factoring in the punting method, it is possible this kick was much more likely to be either muffed or not returned at all. Even if a punter can induce only one extra muff a season, any NFL fan can tell you that a muffed punt can be the difference in a game. Therefore, it may be worthwhile to investigate some sort of statistic that considers how often a punter induces muffs. This may lead to more insight into how NFL teams pay punters than our EPAA method does.

Conclusion

Both general and specific EPAA lend and insight into how effective NFL punters are. They bridge the gap between the academic work of Pasteur et al. (2018), who model nearly every conceivable factor, and other, less formal work which often overlooked important issues regarding return length and penalty yardage. While punting and football statistics in general are complex subjects, our methods lead to a relatively intuitive and understandable statistic that can improve a fan’s experience by giving them a better understanding of the game they enjoy.

References

Author: Chuck Zodda. (n. d.) http://insidethepylon.com/author/chuck-z/

Bien, L. (2018, December 11). This is the golden age of NFL punting. SB Nation. https://www.sbnation.com/2018/12/11/18058128/nfl-punting-art-form-golden-age-evolution.

Bois, J. [SB Nation]. (2019, February 19). *The search for the saddest punt in the world* [video file]. https://www.youtube.com/watch?v=F9H9LwGmc-0

Burke, B. (2014a, December 7). Win probability and win probability added explained. https://www.advancedfootballanalytics.com/index.php/home/stats/stats-explained/win-probability-and-wpa

Burke, B. (2014b, December 7). Tackle factor explained. http://www.advancedfootballanalytics.com/index.php/home/stats/stats-explained/tackle-factor

Chait, J. (2004, December 12). Sabermetrics for football. *The New York Times*, p. 91. https://www.nytimes.com/2004/12/12/magazine/sabermetrics-for-football.html

Costa, G. B., Huber, M. R., & Saccoman, J. T. (2008). *Understanding sabermetrics: An introduction to the science of baseball statistics*. Jefferson, NC: McFarland & Company.

EPA/punt (2020). https://puntalytics.github.io/metrics.html

Gross yards = gross: The Puntalytics primer you didn't know you were waiting for. (2019, September 4). https://www.bleedinggreennation.com/2019/9/4/20850296/gross-yards-gross-the-puntalytics-primer-you-didnt-know-you-were

Holt, R. (1989). *Sport and the British: A modern history*. Clarendon Press.

Kyle, D. (2015). *Sport and spectacle in the ancient world* (2nd ed.). John Wiley and Sons, Inc.

Major League Baseball (n.d.). On-Base percentage. http://m.mlb.com/glossary/standard-stats/on-base-percentage#:~:text=Origin%3A,an%20official%20statistic%20until%201984

nflfastR. (n. d.) https://www.nflfastr.com/

Osborne, J. A., & Levine, R. A. (2017). Shrinkage estimation of NFL field goal success probabilities. *Journal of Sports Analytics*, *3*(2), 129–146. doi: 10.3233/JSA-16140

Pasteur, R. D., Howerton, E., Pozderac, P., Young, S., & Moore, J. (2018). A flight-based metric for evaluating NFL punters. *Journal of Sports Analytics*, *4*(3), 201–213. doi: 10.3233/JSA-180164

Petchesky, B. (2013, August 29). Beware the knuckle punt. https://deadspin.com/beware-the-knuckle-punt-1223314768

Puntalytics [@ThePuntRunts]. (n. d.) https://twitter.com/ThePuntRunts

Romer, D. (2006). Do firms maximize? Evidence from professional football. *Journal of political economy, 114*(2), 340-365. doi:10.1086/501171

SHARP – Scrimmage help/hurt adjusted real punting (2019). https://puntalytics.github.io/metrics.html

Tymins, A. (2014). Assessing the value of NFL punters. *Harvard Sports Analysts Collective*.

Vrentas, J. (2018, January 11). *Punting takes a left turn*. https://www.si.com/nfl/2018/01/11/lefty-left-footed-punters-bill-belichick-dustin-colquitt-zoltan-mesko-ryan-allen.

Zodda, C. (2016a, September 16). Truth in numbers: How punting statistics can lie. http://insidethepylon.com/film-study/film-study-nfl/special-teams-film-study-nfl/2016/09/16/punting-statistics-can-lie/

Zodda, D. (2016b, 22 January). ITP glossary: Target distance punted. http://insidethepylon.com/football-101/glossary-football-101/2016/01/22/itp-glossary-target-distance-punted/

Zodda, D. (2016c, 22 January). ITP glossary: Pin deep punt. http://insidethepylon.com/football-101/glossary-football-101/2016/01/22/itp-glossary-pin-deep-punt/

2019 NFL records and milestones. (2020, 3 January). https://operations.nfl.com/updates/the-game/2019-nfl-records-and-milestones/

2019 punter salaries. (2020). https://overthecap.com/position/punter/2019/

Appendix A

Results

Table A1

2019 Punters Ranked by the Specific Method

Rank	Punter	Specific EPAA	EPAA/punt
1	Way, T.	13.97	0.177
2	Kern, B.	9.93	0.127
3	Dixon, R.	8.71	0.126
4	Huber, K.	7.91	0.105
5	Anger, B.	6.49	0.144
6	Hekker, J.	5.95	0.090
7	Cooke, L.	5.19	0.069
8	Johnston, C.	3.81	0.053
9	Martin, S.	3.39	0.045
10	Morstead, T.	3.21	0.054
11	Gillan, J.	2.22	0.035
12	Wile, M.	2.19	0.365
13	Colquitt, B.	1.89	0.031
14	Bailey, J.	1.41	0.017
15	Winslow, R.	0.78	0.130
16	Dickson, M.	0.67	0.009
17	Bosher, M.	0.33	0.037
18	Allen, R.	-0.44	-0.016
19	Lee, A.	-0.84	-0.014
20	O'Donnell, P.	-1.37	-0.017
21	Long, T.	-1.55	-0.032
22	Sanchez, R.	-1.91	-0.032
23	Haack, M.	-1.92	-0.028
24	Wishnowsky, M.	-2.34	-0.045
25	Edwards, L.	-2.37	-0.027
26	Daniel, T.	-2.38	-0.216
27	Colquitt, D.	-3.16	-0.066
28	Koch, S.	-3.39	-0.085
29	Redfern, K.	-4.59	-0.510
30	Cole, A.	-5.21	-0.078
31	Berry, J.	-5.87	-0.079
32	Scott, J.	-5.89	-0.076
33	Wadman, C.	-8.29	-0.106
34	Pinion, B.	-10.72	-0.188
35	Jones, C.	-10.79	-0.216
36	Palardy, M.	-10.91	-0.145
37	Bojorquez, C.	-14.07	-0.178

Table A2

2019 Punters Ranked by the General Method

Rank	Punter	General EPAA	EPAA/punt
1	Way, T.	14.05	0.178
2	Kern, B.	11.23	0.144
3	Anger, B.	8.99	0.200
4	Dixon, R.	8.83	0.128
5	Huber, K.	7.1	0.898
6	Hekker, J.	6.92	0.105
7	Cooke, L.	6.88	0.092
8	Johnston, C.	5.27	0.073
9	Morstead, T.	4.61	0.077
10	Colquitt, B.	4.59	0.075
11	Martin, S.	4.41	0.058
12	Gillan, J.	4	0.063
13	Bailey, J.	3.39	0.042
14	Wile, M.	1.92	0.320
15	Wishnowsky, M.	1.86	0.036
16	Winslow, R.	1.71	0.285
17	Dickson, M.	0.87	0.012
18	Lee, A.	0.64	0.010
19	Bosher, M.	0.2	0.022
20	Sanchez, R.	-0.06	-0.001
21	O'Donnell, P.	-0.2	-0.002
22	Allen, R.	-0.4	-0.014
23	Long, T.	-0.47	-0.010
24	Haack, M.	-0.9	-0.013
25	Edwards, L.	-0.94	-0.011
26	Daniel, T.	-2.36	-0.215
27	Koch, S.	-2.52	-0.063
28	Colquitt, D.	-2.92	-0.061
29	Cole, A.	-3.54	-0.053
30	Redfern, K.	-4.81	-0.534
31	Scott, J.	-5.66	-0.074
32	Berry, J.	-6.38	-0.086
33	Wadman, C.	-8.85	-0.113
34	Pinion, B.	-9.34	-0.164
35	Jones, C.	-10.34	-0.207
36	Palardy, M.	-11.42	-0.152
37	Bojorquez, C.	-12.61	-0.160