Play Calling in the National Football League: An Analysis

Divya Parmar

Abstract

In the National Football League, events are broken into individual plays called downs. During each possession, a team gets four downs to obtain ten yards to keep possession of the ball. However, if a team does not obtain ten yards within its first three downs, it will kick the ball away on its fourth down. This puts an onus on 3^{rd} down plays. This research looks at play selection during the critical third down. Here we show that running the ball on 3^{rd} down in short yardage (typically defined as three yards or less to go) is optimal, so is passing in 3^{rd}-and-long (typically defined as 11 or more yards to go). This research builds on previous work in regards to play calling, and controls for location on the field. Third down success rate varies significantly across different yards-to-go groupings, but does not differ across yard line on the field where the play took place. With that insight, coaches should optimize strategy for yards to go, but not worry about where the offense is on the field or what sequence the play is. With NFL teams worth billions, and coaching turnover remaining high, coaches cannot afford to ignore data that will lead to more wins.

Introduction

Imagine a chilly Sunday in the fall, as 80,000 screaming fans are in the stands watching their favorite team. The fans are on their feet: it's third down and three yards to go, their offense trying to get a new set of downs to keep possession of the ball. The quarterback is lined up in the shotgun, surveying the defense. He takes the snap, scanning the field, and fires a pass downfield.

Let us pause here. Was that the right decision?

Before we get to that, let me explain American football to provide context for this analysis. In American football, the field is 100 yards long bookended by two 10-yard end zones. Teams score points by either getting the ball into the end zone by pass or run (worth 6 points) or kicking the ball through uprights in end zone (field goal worth 3 points).

Events in football are called plays, signaled by a snap of the ball until the play ends. These are distinct events that behave like a Markov chain. A Markov chain is defined as events such that being in one state, we can assign a probability matrix of moving to another state, regardless of prior history.

In the NFL, we can predict future states such as a touchdown or first down based on the present, even without past history of the previous play. For example, we can put an expected value on possession of the ball with 1st down and 10 yards-to-go on the 30 yard-line, regardless of how the team ended up with the ball at that position.

Markov chain analysis is the foundation of various NFL metrics. For example, Football Outsiders publishes win probability charts that assign each team a win probability at every second of the game. This is only possibly because of historical data, which says that teams with that given score and field position win "x" percent of the time. For example, road team with the ball at its own 20 yard-line down 7-0 with 10 minutes left in the 1st quarter win 40 percent of the time. Extending this knowledge, each play can be assigned a "win probability added" at that given point in the game based on what new "state" the play transitioned the game to.

When a team has the ball, they are considered on offense and have 4 "downs" to gain 10 yards or surrender possession of the ball to the other team. For this reason, teams who do not gain the 10 yards within 3 plays will often punt on 4th down to give the opponent worse field position (or kick a field goal assuming they are in range).

Because 4th downs are essentially an acceptance of the drive ending, 3rd down becomes essential. Gaining the necessary amount of yards and converting the 3rd down keeps the drive going and allows a team the opportunity to score points.

Decision making on 3rd down is key, especially to coaches and player-personnel decision makers. There is a high turnover of NFL coaches, as in a league with 32 franchises, 68 coaches were fired in the past 10 years. The average tenure of an NFL coach is four seasons. In this highly competitive environment, it is vital to use data to gain the extra edge in game situations. As 3rd down is arguably the most important down in the game, one that swings possession and games, coaches are key stakeholders for this insight.

After creating various yards-to-go "bins" and testing for robustness, I found that the optimal in-game strategy is to run the ball on 3rd down and short yardage to go, and to pass the ball on 3rd down with long yardage to go.

Methods

To investigate 3rd down behavior, I obtained play-by-play data from Armchair Analysis[1]; the dataset was every play from the first eight weeks of this NFL season. Since the dataset was clean, and we know that 80 percent of the data analysis process is cleaning, I was able to focus on the essential data manipulation to create the data frames and graphs for my analysis. I used R as my programming language of choice for analysis, as it is open source and has thousands of libraries that allow for vast functionality.

I loaded in my csv file into RStudio[2] (my software for the analysis). First, I wanted to look at offensive drives themselves, so I generated a drive number for each drive and attached it to individual plays dataset. With that, I could see the length of each drive based on the count of each drive number.

Then, I moved on to my main analysis of 3rd down plays. I created a new data frame, which only included 3rd down plays which were a run or pass (excluding field goals, penalties, etc). I added a new categorical column named "Distance," which signified how many yards a team had to go to convert the first down. Using conventional NFL definitions, I decided on this:

Yards to Go	Distance
1-3	Short
4-6	Medium
7-10	Long
11+	Distant

But in order to validate these bins, I also changed the definition slightly to the two groupings below. I tested these groupings as well to see if my analysis reached the same results.

Yards to Go	Distance	Yards to Go	Distance
1-4	Short	1-2	Short
5-8	Medium	3-4	Medium
9-12	Long	5-6	Long
13+	Distant	7+	Distant

I then plotted conversion rate versus number of attempts, essentially an efficiency trade-off chart, for each distance and play type (run or pass) combination. I added vertical confidence bands, derived from the conversion rate (p) and number of attempts (n).

$$standard\ error = \sqrt{\frac{p(1-p)}{n}}$$

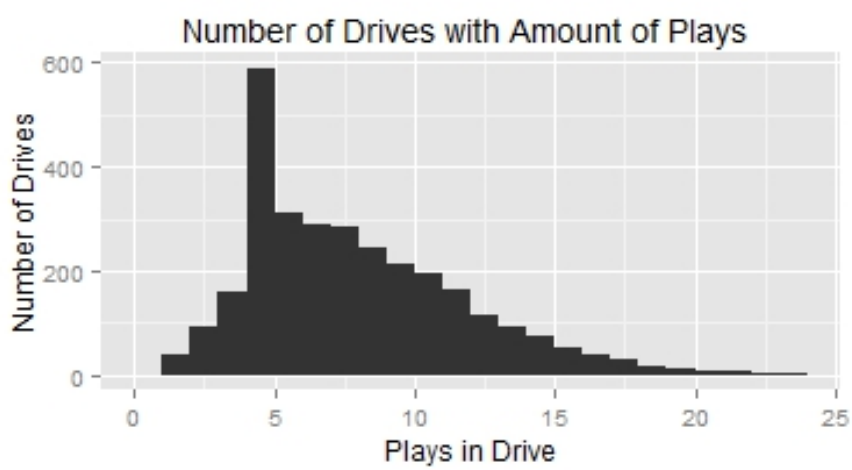
In addition, I investigated a couple more things. I looked at conversion rate on 3rd down based on yards to go until the first down (further would mean lower conversion), and if that conversion was different based on play type. I also looked at location on the field relative to the end zone (measured by the variable yfog aka yards from own goal-line).

I grouped data together based on either yards to go or yard line using the dplyr package, and I used ggplot2 for all my graphs.

Analysis

In looking at offensive drives, I created a drive id number for each drive to allow for deeper analysis. I did this by using a play sequence variable in the data. I looped through the sequence variable, generating a new drive id every time the sequence number went back to 0. With this id and some basic manipulation, I could count how many drives had a given number of plays.

The key for an offense is to keep drives going, and a higher number of plays in the drive is an indicator of success. The importance of 3rd downs is evident, and one would expect that converting 3rd downs would happen very often. But when plotting a histogram of yard per drive [Table 1], most drives were only four plays, which means three plays followed by a punt. Teams were often not picking up a single first down on their drives.



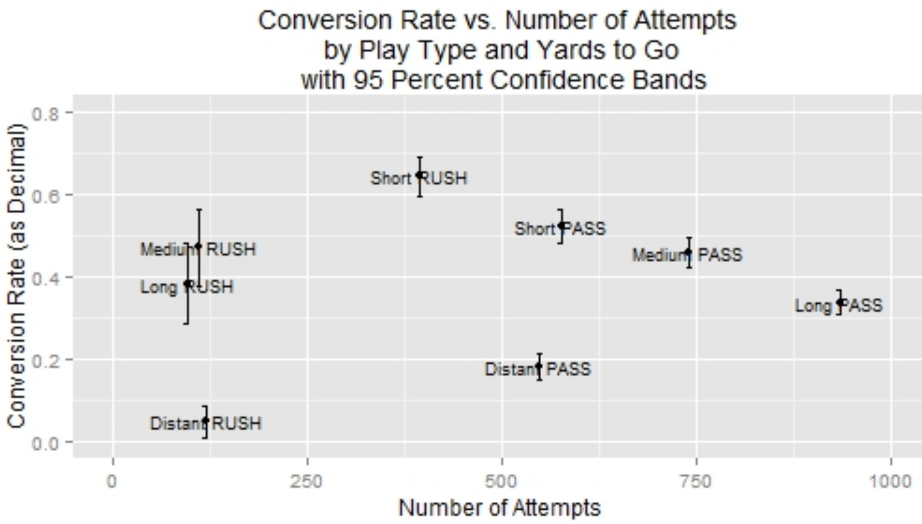
[Table 1]

To assess what factors were important to focus on in terms of third down success, a binary variable, I set up a logistic regression using the same play-by-play data. In looking at the dataset, I excluded during-play variables such as QB sack, tackle, or fumble. In my regression, I predicted obtaining first down on quarter of the game, yards from own goal that play occurred, my “Distance” variable defined previously (as a factor), the play number on the drive, and play type (pass or run, as a factor).

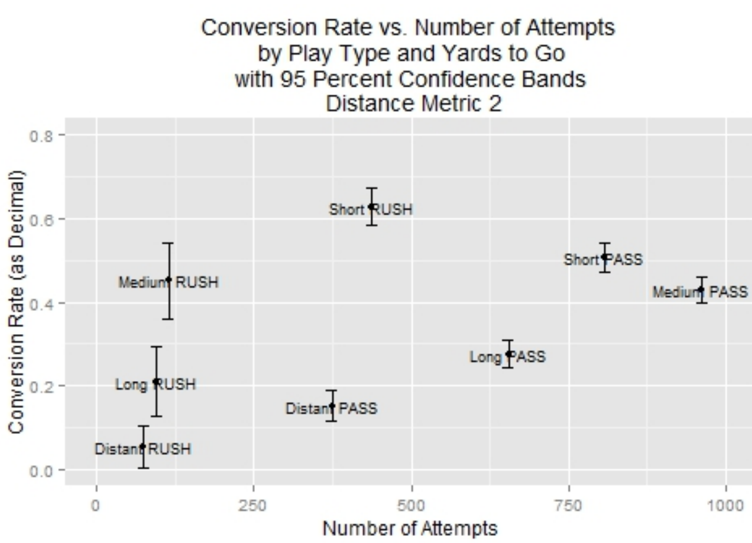
I ran this regression with all three “distance” variables, and the result was the same. Quarter and play sequence were not significantly; the “distance” factors, play type, and yard-line on the field were significant. Because of this, I decided to focus on these three factors in my analysis.

```
Coefficients:
                               Estimate Std. Error z value Pr(>|z|)    
(Intercept)                    0.576086   0.133760   4.307 1.66e-05 ***
qtr                           -0.014225   0.031442  -0.452   0.6510    
yfog                          -0.005164   0.002092  -2.469   0.0136 *  
as.factor(distance)Distant -2.011381   0.129818 -15.494  < 2e-16 ***
as.factor(distance)Long    -0.921134   0.097757  -9.423  < 2e-16 ***
as.factor(distance)Medium  -0.412149   0.098249  -4.195 2.73e-05 ***
dseq                          -0.008105   0.018433  -0.440   0.6602    
typeRUSH                       0.201988   0.093553   2.159   0.0308 *  
---
```

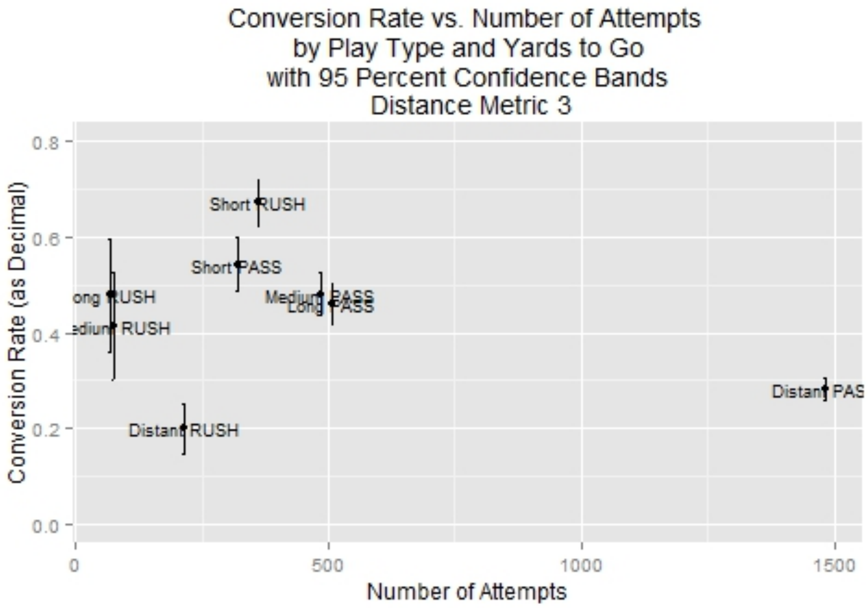
When it comes to 3rd downs, I plotted conversion rate and attempts [Table 2a] for each distance (short, medium, long, distant) and play type (run, pass). I also created the same chart for the two other "distance" bins [Table 2b and 2c].



[Table 2a]



[Table 2b]

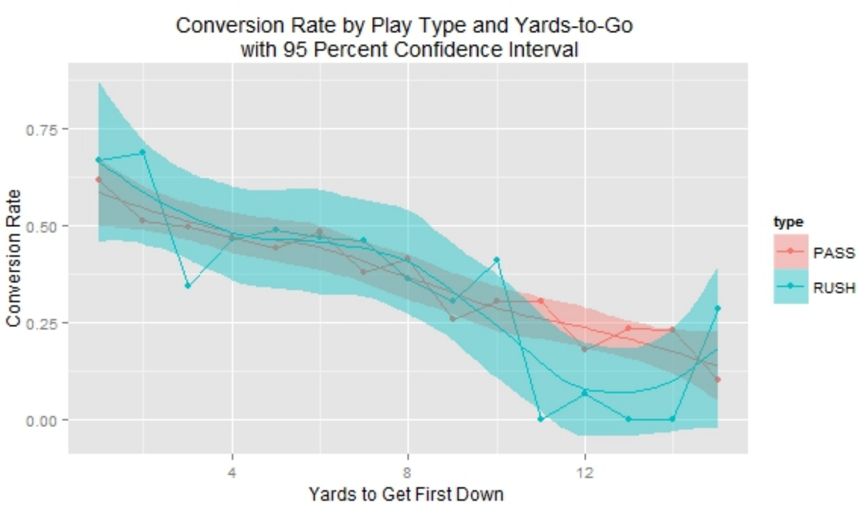


[Table 2c]

I then compared passing and rushing for each distance, using a two sample z-test for proportions to see if there is a statistically significant difference between the two play types.

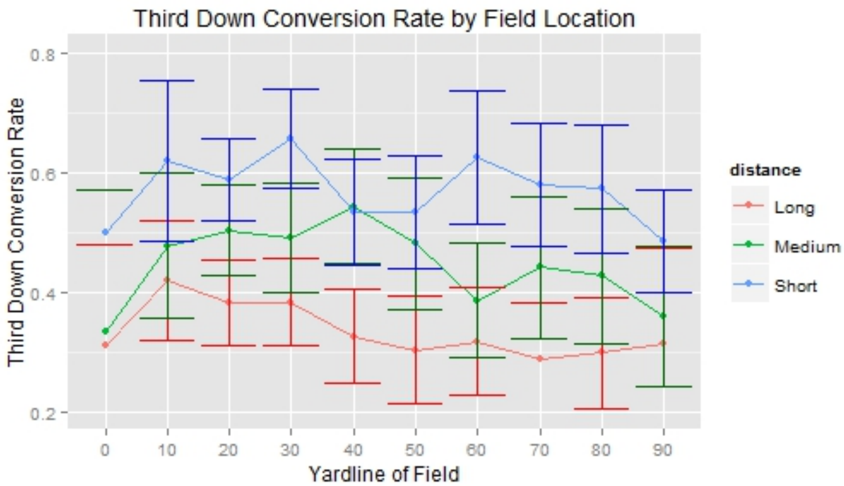
I found that rushing was significantly better in "short" situations, passing was significantly better in "distant" situations, and there was no significance in "medium" and "long" situations[3]. To build on the distances, I looked at individual yards-to-go amounts and plotted the conversion rate with a 95 percent confidence band [Table 3]. It validated my previous work and showed play type only mattered for short and distant.

To further validate these conclusions, all three "distance" groupings followed the same significance patterns.



[Table 3]

I also looked at location on the field, measured as yards from own goal-line [Table 4]. I plotted the conversion rate by 10 yard intervals, adding 95 percent confidence bands for short/medium/long distance conversion rates. In looking at the graph, there is no statistically significant difference in conversion rate for any of the three situations across the yard line intervals.



[Table 4]

Conclusion

In conclusion, there are a few recommendations to potential clients, if you're listening (San Francisco 49ers, are you there?). Running the ball on 3rd down in short yardage (typically defined as three yards or less to go) is optimal, so is passing in 3rd-and-long (typically defined as 11 or more yards to go). Note, this data confirms what the eye test already tells us. Furthermore, 3rd down ability does not change across the field, so don't be overconfident if you're in opponent territory.

This paper focuses on a critical subset of decisions in the NFL, 3rd down. It adds to the discussion around play-calling and gives coaches data to back up their intuition. However, it shouldn't be taken as a blanket answer for all situations, as individual matchups may present different outcomes.

The Markov application to the NFL's discrete event structure fits nicely, and this is the foundation for game prediction and gambling lines. Thus, similar research has been done on this topic[4, 5]. However, in terms of future research, the next step would be to look at specific types of running and passing plays and their success rate. That would give further insight to coaches and be more actionable.

Appendix

1. The data was downloaded here: http://armchairanalysis.com/data.php.

2. The software and R language can be downloaded here: https://www.rstudio.com/.

3. This below is the code for the significance tests for proportion.

```
> #Distant
> prop.test(conv_rate$converts[1:2],conv_rate$numberattempts[1:2],alternative
="greater")

        2-sample test for equality of proportions with continuity correction

data:  conv_rate$converts[1:2] out of conv_rate$numberattempts[1:2]
X-squared = 11.9252, df = 1, p-value = 0.0002769
alternative hypothesis: greater
95 percent confidence interval:
 0.08458557 1.00000000
sample estimates:
   prop 1    prop 2
0.1821494 0.0500000

> #Long
> prop.test(conv_rate$converts[3:4],conv_rate$numberattempts[3:4],alternative
="two.sided")

        2-sample test for equality of proportions with continuity correction

data:  conv_rate$converts[3:4] out of conv_rate$numberattempts[3:4]
X-squared = 0.6259, df = 1, p-value = 0.4289
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.15374310  0.06171062
sample estimates:
   prop 1    prop 2
0.3394004 0.3854167

> #Medium
> prop.test(conv_rate$converts[5:6],conv_rate$numberattempts[5:6],alternative
="two.sided")

        2-sample test for equality of proportions with continuity correction

data:  conv_rate$converts[5:6] out of conv_rate$numberattempts[5:6]
X-squared = 0.0167, df = 1, p-value = 0.8971
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.11600382  0.09265217
sample estimates:
   prop 1    prop 2
0.4615385 0.4732143

> #Short
> prop.test(conv_rate$converts[7:8],conv_rate$numberattempts[7:8],alternative
="less")

        2-sample test for equality of proportions with continuity correction

data:  conv_rate$converts[7:8] out of conv_rate$numberattempts[7:8]
X-squared = 13.9871, df = 1, p-value = 9.204e-05
alternative hypothesis: less
95 percent confidence interval:
 -1.00000000 -0.06852975
sample estimates:
   prop 1    prop 2
0.5225694 0.6455696
```

4. Markov analysis of NFL overtime: http://harvardsportsanalysis.org/2014/01/modeling-nfl-overtime-as-a-markov-chain/.

5. Football Outsiders' analysis: http://www.footballoutsiders.com/info/methods.