2016

The Drivers of Success in the NFL: Differences in Factors Affecting the Probability of Winning Based on First Half Performance

Zach Alleva
Claremont McKenna College

Claremont McKenna College

The Drivers of Success in the NFL: Differences in Factors Affecting the Probability of Winning Based on First Half Performance

Submitted to
Professor Darren Filson

by
Zach Alleva

for
Senior Thesis
Spring 2016
April 25, 2016

Acknowledgements

This thesis would not have been possible without the helpful advice of both Professors Filson and Weidenmeir. Their advice in refining my direction of study and the content of my work was invaluable. I would also like to thank my friends and family for their continuous support. Finally, I must thank my fellow Ryal Lab seniors.

Abstract

This paper explores how changing various end game statistics effects a given teams probability of winning a game in the National Football League (NFL). Data from the 2000-2016 NFL seasons is split into two subsets, one for teams winning at halftime, another for losing teams. Using this data an empirical model is estimated to study how the determinants of a team's success differ between the two sets of data. Overall, the factors which determine a team's outcome are consistent between the two subsets, varying primarily by magnitude of the effect.

Table of Contents

I. INTRODUCTION

In the last decade the NFL has become a behemoth among sports leagues worldwide. In 2015, the league earned in excess of $12 billion in revenue, with profits of about $1 billion (Isidore). This figure only looks to be climbing, as is the league popularity. The most recent championship, Super Bowl 50, reached over 100 million viewers. This count did not include those who watched the game online (Pallotta and Stetler). It is fair to say that the NFL is an extremely successful American institution.

The success the NFL enjoys has resulted in much "armchair coaching", where, like with any popular sport, the fans and sportswriters attempt to offer their best explanations for why a team is successful on any given Sunday. Most of these explanations are grounded in nothing more than conjecture about what seems important, or are backed up by a vague explanation of trends they appear to see.

The goal of this study is to empirically investigate the drivers of success in the NFL in order to understand how these drivers affect a team's probability of winning. I am particularly interested in analyzing any differences there might be between what factors best predict success for teams which are ahead or behind, respectively, at halftime. The reason for this methodology is the belief that teams may operate differently depending on their status at halftime. It may be that teams which are behind throw the ball significantly more than teams which are ahead. According to conventional wisdom, teams which are ahead are said to try and run out the clock and run the ball more. It is notions like these which I will analyze throughout this paper.

I approach the question by dividing my data into two subsets; teams which are ahead at half and teams which are behind. This allows me to analyze the effects of various statistics on each group of data and compare the results. The data consists of various end-game statistics,

such as time of possession, completion percentage, interceptions and passing yards, among others. These are further divided into first and second half statistics. This allows me to analyze the effects of each factor separately by half so it is clearer how the variable affects the outcome. The data covers the 2000-2016 NFL seasons, a total of 16 seasons and 3908 games.

In general it appears that the most significant factors which affect a team's probability of winning a game or similar whether a team is leading or trailing at halftime. Though there are some marginal differences as to what factors are particularly important. This could imply that there are core strategies which are constant despite a team's position. This is only natural, since certain factors such as passing and first downs will be necessary for a team to move down the field and score. The differences in factors show where strategies might differ between leading and trailing teams.

There were a few issues present regarding lack of complete data. In some cases data was only available for specific seasons, in other cases data which would have been useful in my analysis was unobtainable with the techniques known by me. However as whole the analysis presents useful insights pertaining to the drivers of success in the NFL.

II. LITERATURE REVIEW

The field of sports has long been an area of interest to fans and managers alike. The allure of having the ability to accurately assess the value of various players or schemes understandably draws the attention of those with a vested interest in a team's performance. The goal of any statistical analysis in sports is to explain the relative value of that which produces the statistic.

Sports analytics first began to be used in earnest during the early 1990's with the advent of saber metrics and its utilization by Major League Baseball teams. Proponents argued that the then-current statistics used to judge players were out of touch with how the games were actually lost or won (Lewis, 27). It was time for a new era of sports analytics to be born. Though adoption of advanced analytics in the NFL has lagged behind the MLB, some academics and fan-driven studies have begun to focus on the variables present in football games and created some very in depth studies… One area of economic interest is creating an efficient team because of salary cap restrictions, as opposed to baseball where they have small/large market issues.

Few studies have seriously considered how best to determine the likelihood of a team's victory in the NFL. The majority of studies that have been completed look primarily at a team's overall record or use post-game statistics to determine who the likely winner of a game was. One of the earliest and most common ways to do so involves using the Vegas spread to predict the winner of regular season games (Stern, 1991). This study found that the Vegas point spread (which is around 66% accurate) was a viable way to predict a team's probability of winning.

One of the first questions anyone asks about football is whether offense or defense is more important to winning. A study by Robst, VanGilder, Berri and Vance (2011studied the probability of a team winning a game in the postseason based on how well their offense or defense performed over the course of the regular season, to help and NFL franchise better

allocate salary cap money. The findings suggest that teams which are better on both offense and defense than their opponents are more likely to win. However they found that neither a better offense nor a better defense alone contributed to a significantly higher probability of winning. In fact the benefit of having a great offense and poor defense and vice versa were statistically similar. The results indicate that defense and offense are equally beneficial. Additionally they found there to be a significant effect for the phenomenon home field advantage, where home teams are more likely to win.

Another group of researchers, Hadley, Poitras, Ruggiero, and Knowles (2000) similarly looked at various post-game statistics during the regular season from 1969-1993 with the goal of measuring the performance of NFL teams and coaches. By estimating a production frontier for the number of wins for a given team based upon offensive and defensive performance statistics, they could examine the efficiency of an NFL team relative to their potential. They found that better offensive and defensive statistics meant a team had a higher probability of winning and should win more games per season. They also looked at coaching efficiency and found that a better coach can contribute three or four wins above average per season. In addition, coaches with more experience tend to be more efficient. Hofler and Payne (1996), also examined team efficiency in the NFL by looking at how close teams played to their offensive potential. They found that teams were extremely efficient compared to a stochastic production frontier modeling their potential. Offenses scored within one point on average of what the model estimated their potential to be. This shows that coaches understand how to effectively utilize the talents of their players.

Atkinson, Stanley, and Tschirhart (1988) also examine the ability of various post-game statistics in order to predict outcomes for regular season NFL games. The results hold in line

with the previous studies, finding that teams with better offensive and defensive statistics are more likely to be victorious. For example, teams with more rushing and passing yards, fewer turnovers, and fewer sacks are more likely to win. On the defensive side, teams that allow fewer rushing and passing yards, more sacks, and force more turnovers are also more likely to win. It is interesting to note that their estimate for coaching ability was not significant, unlike previous studies; however they based the indicator off whether the coach had won coach of the year in years prior. This may not be an accurate representation of coaching ability as only one coach a year can win and sometimes voting awards have a political nature to them as opposed to purely statistic.

In order to control for different strategies in the second half based on how the first half went Arkes (2011) uses solely first half statistics to predict outcomes. The reasoning is that often in the second half teams will alter their strategy to maintain a lead; the goal of the study is to determine which strategies are beneficial to gaining a lead in the first place. The data comes from the 2005 NFL season; the author finds that a first-half passing advantage significantly increases the probability of winning, while a rushing advantage in the first half does not. However a rushing advantage using full-game statistics increases the probability of winning, this stems from a change in the leading teams strategy during the second half. This implies that the passing game seems more important than rushing for a team's victory, however a strong rush game can be helpful in maintaining a lead. In addition, Arkes found that penalty yardage was not significant in determining the eventual victor. Consistent with earlier studies, turnovers and sacks negatively affect a team's chances of winning. Another facet of the game he focused on was time of possession; it is typically believed that the team which controls the ball more has a greater

chance of winning. The research found that the time of possession variable is not significant to determining the outcome of games.

Other researchers, Pitts (2016), have looked into the value of previous playoff experience to see whether more experienced teams are better in the post season. Using data from the 1966 to the 2012 NFL postseasons, the hypothesis was that teams with more veteran players and experience would perform better in the playoffs because they were used to the pressure. However, the results showed that overall player experience had no significant effect on a team's success in the playoffs. Interestingly, for specific positions such as quarterback there is a larger effect. New quarterbacks are about 14% less likely to win, whereas there is no effect, positive or negative, after the first year. In accordance with Robst et al. (2011), Pitt finds that offensive and defensive marginal production are both positive and statistically similar, making it difficult to determine if one is of more importance. In line with Arkes' (2011) findings relating the importance of rushing and passing, Pitts shows that only passing offense productivity is a significant determinant of post season outcomes. The results provides more evidence that passing is the key to victory for NFL offenses and suggests that teams should focus salary cap allocation towards pass defense and away from the running game. Some other variables of note include the finding that teams which go on hot streaks at the end of season do not perform any better than those who have been losing, in addition home field advantage is still found to be advantageous and the team with the better regular season record is more likely to win.

McGee, Van Scyoc, Burnett (2012) look at salary cap allocation on the offensive and defensive side of the ball in order to determine the optimal salary cap allocation. Using data from 2000-2009 they find that teams which focus on one side of the ball over the other tend to perform

better. This is interesting because it seems to confirm, from another perspective, that offense and defense seem to contribute almost equally to a team's success.

Other researchers have focused on the ability to predict plays in the NFL. Burton and Dickey (2015) have created a statistical model to predict whether a passing or running play will be called at any point during a game. Using play by play data from 2011-2014 they use inputs such as field position, time of possession and others to determine the likelihood of the type of play call. In testing 20 randomly selected games, they successfully predicted the play 75% of the time. This has implication for attempting to predict the winning team, possibly by using the model to see if the winning team's plays were predicted correctly more often than the losing teams.

III. HYPOTHESIS DEVELOPMENT

Initially, I was interested in simply understanding what the relative effects of various end-of-game statistics were on the outcome of an NFL game. Much of the previous literature is focused on looking at full game statistics. However, after reading Arkes 2011 paper on the effects of first half statistics on outcomes I became interested in studying the second half of NFL games specifically. I reasoned that since the team which holds the lead at the half wins about 77% of the time it would be worthwhile to conduct a study on how teams maintain the lead they acquired in the first half. The goal of this study is to understand the drivers of success in the NFL, with a focus on how those drivers evolve throughout the course of a game between the first and second halves. This is interesting because it allows for the breakdown of games into several components which can then be analyzed and the information gathered from this could be used to isolate specific areas teams need to improve upon in order to succeed in different game situations.

The general theory behind the idea is that a team's game strategy and therefore their statistics, which are explanatory of the teams strategy, may be different in the first and second half depending on whether they hold a lead going into halftime. For example, it is widely believed that winning teams change to a strategy of delay in order to wind down the clock, while the losing team will become more aggressive in order to catch up (Arkes, 2011). One way teams might do this is to run the ball more in the second half. By giving a quick look at the summary statistics for teams which are in the lead at halftime, we see that the average percentage of pass plays declines by nearly 10% in the second half if a team holds the lead, from 55% to 46%. While, conversely, teams which are behind increase their pass plays by slightly less than 10% on average (Tables2 and 3).

An alternative explanation for this increase in the percentage of run plays could be that leading teams are afraid of turning the ball over, another important part of football games. Again, looking at just the summary statistics, we see that teams are less likely to recover a fumble than have the ball intercepted. In other words, it is riskier to run a pass play than a run play if a team is worried about losing possession of the ball. Additionally, it seems likely that turnovers may be more important in the second half than in the first half. This is because, if the previously stated theory of delay is true, then every extra possession will be particularly important, especially in the second half.

Given this information it seems worthwhile to study the effects of various post-game statistics on the outcome of NFL games. The overarching goal of this study is to understand which statistics are the best predictors of success in each half and to understand if teams which have a lead going into the half have different predictors of victory than teams which are trailing. I have two subsets of data, each of which I presume, will have different variables which will best predict the outcome.

1) For the set of games where the team is leading at half that the probability of a team winning will be dependent on strong offensive measures in the first half, such as passing yards, number of first downs and rushing yards. In the second half I expect a low percentage of pass plays, higher rush yards, a low number of turnovers against and a high number for to be important. I also expect time of possession in the second half to be an important indicator of victory. In addition it is unlikely teams in the lead will have rookie quarterbacks, as rookie quarterbacks only win 36% of games they play.

2) For the set of games where the team is behind at half I expect the results to be somewhat inversely related to the previous set. While I do expect the trailing team to increase the percentage of pass plays, I doubt the effect will be as important as other variables for predicting success, since all losing teams will do it, not only the ones who win. However I do expect it to be significant in the regressions. It is also likely that turnovers will become especially important, as the trailing team will have to play an impeccable second half in order to regain the lead. Similarly, completion percentage will be important, as they will need to play very efficiently to maximize their plays and time of possession if in fact the opposing team is stalling. Starting field position will also be an important variable, the closer the team is to the goal line the easier their drive will be and it will take less time as the leading team tries to run out the clock.

With this framework I will attempt to analyze how the differences in predictors of success between leading and trailing teams in the second half and how these statistics affect their probability of winning the game. A number of other variables are included which help to provide a more robust model of an NFL game and which may also have an interesting effect on the outcome.

IV. DATA & VARIABLES

The data for this study was initially obtained from arcmchairanalysis.com. A site devoted to providing easily accessible, advanced and thorough data on NFL games. The data covers the 2000-2016 seasons, 16 seasons in total, consisting of 3908 individual games. I was unable to gather any earlier data because 2000 is the earliest season armchairanalysis.com has data available for. Armchairanalysis.com obtains their data directly from the NFL and is guaranteed to be accurate, given the data from the NFL has been recorded correctly.

The dataset was initially comprised of multiple files, each containing statistics on different aspects of NFL games, such as play-by-play data, drive-by-drive statistics, full game statistics or player-specific statistics. By combing through, cleaning up and consolidating the data I was able to produce two sets of data for each game, one for each team, as well as separate statistics for the first and second halves in order to tests my hypotheses. All variables relate to a given teams performance in a specific game. Most of the measures used are ex-post as opposed to a priori, so they are only known after the completion of the game. While this is means the results are not as useful for predicting the outcomes of games, they can still be analyzed to understand the relative impact of different parts of the game on the outcome. This analysis could subsequently be used in order to understand which aspects of the game a team should place emphasis on in the first and second halves.

In addition, as Arkes (2011) mentions, there may be endogeneity issues with statistics observed ex-post. However, I do not believe this is an issue for my research as it is exactly this endogeneity which I am interested in. Arkes believed that second half statistics may skew overall game statistics because teams may change their strategy, and thus their distribution of statistics, in the second half, depending on the results of the first half. Consequentially, he focused

primarily on first half results. In my data I divide the statistics by halves so I can analyze the difference in first and second half statistics to see if the significantly explanatory variables do in fact change between the halves. This may signal that teams change parts of their game plan to account for what happened in the first half. There could also be other explanations for why the important predictor variables change, such as players getting tired from being involved in too many plays in the first half. Primarily, my goal is to determine what it is that successful teams do in the first and second halves of NFL games. Thus, the concern about endogeneity is not relevant to my analysis.

One issue presented by my data is the presence of teams tied at half and at the end of the game. While ties are very rare in the NFL, these teams were still removed from the analysis so as to avoid any possible issues that might arise during the analysis. This removed 694 observations.

Additionally, it would be helpful to have a more complete set of data describing the game. For example red zone data on plays within 20 yards of the opposing team's end zone, was not available for a number of seasons. This could have been helpful for my analysis to look at efficiency in the red zone and could have potentially been more explanatory than raw yardage, because a team could be great at moving the ball down the field but have difficulty converting their success moving the ball to points once they get towards the end zone and the field becomes confined. A few other potentially interesting statistics to look at would be the effect of injuries, types of defensive schemes and team salary cap allocation.

My data is divided into two subsets, one for teams which are ahead at half and a second for teams which are behind. The descriptions of each variable and how they were computed can be found in Table 1. The summary statistics for the variables included in my analysis can be

found in Tables 2 and 3. Following, I provide the motivating factors behind including each variable.

A. Control Variables

In my analysis I control for home field advantage and a teams win or loss streak going into the game. In previous research, Robst et. Al. (2011) and Pitts (2016), a home field advantage has been found to be a significant determinant of victory. Presumably this is because the team gains some sort of advantage, perhaps because their opponents had to travel, from playing at home. As a result, I include it as a measure of a team's performance helps to explain some of the results not captured by my hypothesis.

The team's win/loss streak is included as a measure of the team's recent success and overall strength. This variable is meant to control for the relative strength of two teams. While this is not the best measure, ideally some type of win/loss record adjusted for strength of schedule could be used, it still captures some of the variability in the relative strength of teams; the best teams will have longer win streaks while the worst teams will have longer loss streaks. This helps to avoid issues where a team's overall strength affects other variables, for example a worse team might consistently have lower offensive totals and the streak statistic helps to account for such effects.

The other variables considered are all included in order to study their effect on a team's probability of winning using either; the full set of games, the subset where teams were behind at half, or the subset where teams are leading at half. As I will explain in detail later, I include these three different subsets in order to first, look at full game statistics and then study the differences between teams which lead at half and are behind at half in order to provide understanding of

what the best determinants of success are in an NFL game, depending on how the team is positioned at half. The variables are broken up into five broad groups.

B. Independent Variables

1. Rookies

I include the number of offensive and defensive rookies who play in each game, as well as a binary variable for whether a team has a rookie quarterback or not. The theory is that teams with rookies will not perform at the same level as veterans, especially in a comeback situation, primarily because they are inexperienced and the transition from college to the NFL can be disorienting. This should be particularly true for quarterbacks, who face a much higher level of play with far more athletic and complicated defenses than they faced in college. The remaining variables are all separated by the first and second half, so there are two sets of each.

2. Offense

Six variables are used as measures of a team's offensive success or style. The number of first downs a team earns is included. This is an important measure of a team's offensive success. Unless the team consistently gets big plays, which are relatively rare, dependably earning first downs is the only way to continue a drive downfield and give the team a chance to score. Earning first downs also prevents the other team from getting the ball back and having the chance to score. Thus it is very likely that an increase in first downs will be associated with an increased probability of winning. I also measure the number of times a team goes for it on fourth down. The conventional wisdom is to punt the ball, recently though, outsiders (such as bloggers and sportswriters, less so coaches) have begun to vouch for the numbers behind going for it on

fourth down. I include this variable in order to test the credence of this theory; however there are some issues inherent in the analysis. Most teams do not attempt fourth down plays consistently, so they often take place only in high pressure situations and may not be regularly practiced, thus teams may be less effective at converting on fourth than they possibly could be. Additionally, most of the teams which go for it on fourth down are trailing, which may be an indicator that they are the worse team and would presumably have a more difficult time converting to first down. As we see in Table 6 most fourth down attempts take place by trailing teams in the second half.

A team's total number of rush yards is included in order to study the difference in the importance of rushing effectiveness in the first and second halves. Since we believe rushing is used as a delay tactic in the second half, teams in the lead will likely have a higher number of rush yards. A similar variable used is passing yards. Presumably teams with higher passing yards will be more successful since they are better at moving downfield towards the end zone. Furthermore, it is likely this statistic will be high in the first half for winning teams and high in the second half for trailing teams.

The measure of a team's completion percentage is included because it provides another way to look at a team's passing effectiveness besides using total yards. This could be useful to account for teams which do not throw for a lot of yards but rather, consistently make completions and play safe. Additionally I include the number of two point conversions a team attempts. I expect this number to increase in the second half for teams which are trailing. While most teams eschew going for two points and settle for an extra point, trailing teams will likely need to make riskier plays in order to catch up.

The last variable used to measure a team's offensive effectiveness is time of possession. I include this variable because it is generally believed that winning the time of possession battle will help propel a team to victory. Though Arkes' 2011 study found the effect of time of possession to be insignificant I think it is worthwhile to include, as the sample used in that study consisted of a singular season and Arkes confined his analysis to the post-season.

Also included in the regressions are variables used to study different styles of offense. I include the number of plays in which a team runs a no-huddle offense. Presumably, teams which are behind are far more likely to run the no-huddle offense because they need to save time. It is also possible that teams which can run the no-huddle effectively will gain an advantage by preventing the opposing defense from regrouping between plays. The percentage of pass plays is also included in my analysis. This is used to measure the team's pass/rush play mix. It would be expected that the percentage of pass plays will decrease in the second half for teams in the lead, while the trailing team will step up the number of pass plays called and avoid running the ball.

3. Defense

Three variables are used to account for the impact of a team's defense on the probability of victory. The first is the number of sacks a team's defense earns. The idea is that this measures how much pressure the defense puts on the quarterback and is widely recognized as a way to differentiate pass rush defenses. Teams with a higher number of sacks should be more successful because it causes the opposing team to move farther back from the first down line and thus makes earning first downs, which are relatively important, much more difficult.

The second and third defensive variables are the number of interceptions and fumbles recoveries by the teams defense, both relate to turnovers. Theoretically, team's which can

successfully take the ball away from their opponent should have an increased chance of winning. When teams force turnovers they typically receive the ball with better than average field position, since the opposing team does not have the opportunity to punt the ball away and thus will have a better chance at scoring.

4. Special Teams and Penalties

The final set of variables includes the effects of special teams and penalties. As mentioned above, the field position turnovers create is presumably valuable to the team forcing the turnover. Extending this idea, I include a variable which measures the team's average starting field position. This variable is an amalgamation of the effectiveness of a team's kick return unit, as well as some effect for turnovers, which indicates where on the field the team starts their drive sequence. This variable is included to account for how effective the team is at returning kicks and setting themselves up for a successful drive with good field position. The motivating idea is that, the closer a team starts their drive to the opposing team's end zone, the more success they will enjoy.

Finally, I account for the team's net penalty yardage. Penalty yards a somewhat like free plays in that the team doesn't have to work for them, rather the opposing team just has to make a mistake. Similarly, negative penalty yards are costly, as they can sometimes cost a down and push the team farther back from the first down line. Thus, it would be expected that teams which have higher positive net penalty yards will have a better chance at winning.

V. EMPIRICAL METHODOLOGY

The model employed in the study is a probit regression model, which is estimated using the standard maximum log likelihood procedure, to test for the effect of various post-game statistics on the likelihood of a team winning an NFL game. Since I am regressing on a binary dependent variable (winning/not winning), I utilize the probit model in order to test the hypothesis that the style of gameplay which contributes to a victory changes between the first and second half, dependent on whether a team is leading or trailing at halftime.

Furthermore, I will be utilizing a dprobit function in order to obtain the marginal effects of each variable on the likelihood of winning. This allows me to interpret the effects of individual variables, holding the rest of the variables constant, and compare their relative effects on the probability of winning.

I include one model run on the two subsets of data, for a total of two regressions. The application of this model will allow me to interpret the effects of the individual variables in each half.

$$\text{Prob}(Win) = \beta_0 + \beta_1 home + \beta_2 streak + \beta_3 qbrookie + \beta_4 offrookies + \beta_5 defrookies + \beta_6 sackfor1 + \beta_7 sackfor2 + \beta_8 intsfor1 + \beta_9 \text{intsfor2} + \beta_{10} fumrec1 + \beta_{11} fumrec2 + \beta_{12} startfieldpos1 + \beta_{13} startfieldpos2 + \beta_{14} fd1 + \beta_{15} fd2 + \beta_{16} percentpassplay1 + \beta_{17} percentpassplay2 + \beta_{18} ry1 + \beta_{19} ry2 + \beta_{20} py1 + \beta_{21} py2 + \beta_{22} comp1 + \beta_{23} comp2 + \beta_{24} nh1 + \beta_{25} nh2 + \beta_{26} top1 + \beta_{27} top2 + \beta_{28} netpenydg1 + \beta_{29} netpenyd2 + \beta_{30} go4th1 + \beta_{31} go4th2 + \beta_{32} conv1 + \beta_{33} conv2 + \varepsilon$$

In the model above, the probability of a team winning (Prob(*Win*)) can be explained by the team's in-game statistics, as explained in the previous section.

VI. RESULTS & DISCUSSION

Many of the factors included in my model were found to have a significant effect on the probability of a team winning. For the full results of the probit model, refer to Table 4. Some of the results are expected, while others were somewhat surprising.

Overall, both models are fairly good predictors of the probability that a team will win. For the regression run on the set of games played where the team held a lead at halftime, a total of fifteen out of the thirty-three variables tested were found to be significant at the 5% level or lower, including both control variables. This seems to explain the relatively high pseudo R^2 value of 0.5436, which means the regression has strong interpretative power. Though I should note that there is more difficulty in interpreting the meaning of a pseudo R^2 than the traditional R^2 (Hu, Shao, Palta 2006).

When the model was regressed on teams which were behind going into halftime, fourteen out of the thirty-three variables were significant at 5% or lower. The pseudo R^2 was 0.5494, which is nearly equivalent to the previous regression.

Generally, the results are fairly consistent across both regressions. First I will go through the regressions and discuss the variables which were significant in both subsets of data. Then I will cover the variables of interest which differed between datasets.

A. Variables with Significance in Both Sets

A note of importance about the analysis, every variable which was found to be significant across both data sets occurred in the second half of the game, ignoring the control variables which are not confined to one half. Additionally, only three of these variables had significantly different values in each regression, based upon a chi-squared test. These three variables were

passing yards, completion percentage, and the percentage of plays which were passes, all occurring in the second half. The rest of the coefficients in this section were functionally similar between teams leading and trailing after the first half.

Both of the control variables, home and streak, were found to be highly significant at the 1% level. This verifies the work of previous authors who found there to be a significant effect for home field advantage. Additionally, the marginal effects of the variables in each regression were nearly identical, at approximately 0.05 and 0.007 for home and streak, respectively. According to the results, if a team is playing at home they will be approximately 5% more likely to win the game. Similarly, for every extra game a team has won in a row; they will be 0.7% more likely to win the game. These results are very much expected, teams are typically believed to play better at home since they seem to gain some sort of advantage from having a friendly crowd, plus not needing to travel which can be tiring. It is also expected that teams with longer win streaks are likely to win because generally, the better team will have a longer win streak. Of course if this were some metric like an adjusted strength of schedule, I would imagine it would have more predictive power, since teams may have a longer or shorter streak simply because they had a very easy or very difficult schedule.

The three defensive measures of sacks, interceptions and fumble recoveries were all significant at the 1% level. However, the interesting part is that they were only significant in both regressions during the second half. Sacks had the lowest marginal effect on the probability of victory (3% increase per each sack), though the marginal effect of sacks was not significantly different from the effect of fumble recoveries (5% per fumble recovery). Of the three, interceptions had the highest marginal effect, with each interception increasing the probability of victory by 7% per each interception. This was significantly different from both the effect of

sacks and fumble recoveries. It seems clear that interceptions are the most impactful indicator of a team's success out of the three defensive measures included in the analysis. A possible explanation for this result is that defensive players tend to pick up more yards after interceptions than fumbles because they are usually in one-on-one coverage, while fumbles tend to happen during running plays, where it is much more crowded. Of course more research would be needed to test this theory. The reason interceptions are more valuable than sacks is most likely because it creates a turnover as opposed to a loss of yardage for the opposing team. It is still difficult to explain why these three measures are all significant in the second half as opposed to the first. The most probably explanation is that team's turnovers become particularly important in the second half, regardless of whether a team is leading or trailing because they allow a winning team to cement the victory with an extra possession or conversely, allow the trailing team to catch up.

The team's average starting field position in the second half is another variable which is significant in both regressions. Again, the coefficients are not significantly different and the marginal effect is rather small at only 0.3% increased chance of winning for each extra yard of field position. A better average starting field positon would seem to benefit a team because it puts them closer to the opposing team's end zone, which makes it easier to score. It is difficult to explain why this variable is a positive, nearly identical predictor of victory for both datasets. Initially I had thought it would only matter for trailing teams, as leading teams have less need to score, however it seems good field position benefits both teams equally in the second half.

The statistic measuring time of possession in the second half was found to be a significant predictor of victory for both trailing and leading teams; it was also significantly different between leading and trailing teams. For the leading team a 1 minute increase in time of

possession resulted in a .7% increase in the chance of winning. Interestingly, time of possession had over twice the effect for trailing teams, at a 1.7% increase per minute. The most probable explanation is that in order for teams to score, they need to have possession of the ball, the trailing team needs to score more and more time with the ball gives them that, while preventing the opposing team from lengthening their lead.

The three measures of offense which are significant across both sets are passing yards in the second half, the team's completion percentage and the percentage of pass plays called. For passing yards, the difference between leading and trailing teams is not significant; however for completion and pass play percentage, the difference is significant. The marginal effect for passing yards is 0.11% increase in the chance of victory per every yard thrown. I would have expected passing yards to be more important for trailing teams, which it is slightly, but not at a significant level. So it seems teams benefit equally from increases in passing yards.

For teams with the lead, a 1% increase in pass completion increases the probability of winning by 0.16%. Comparatively for trailing teams, the increased probability of victory per percent increase in pass completion is 0.33%. As predicted, completion percentage is more important for the trailing team, this is likely because they simply need to play a better game in order to catch up.

The marginal effect for a 1% increase in pass plays for teams leading at half is a 1.17% decrease in the likelihood of a win; compared to a 0.8% decrease in probability of victory per 1% increase in pass plays for trailing teams. The difference between these two measures is significant and yields more interesting analysis than previous results. The probable explanation for teams winning at halftime is more straightforward, it seems likely that throwing more passes increases the chances of an interception, which we know has a significant effect on the outcome

of the game. Additionally passes take less time off the clock because if they are incomplete the clock stops, while even a failed run play will keep the clock moving and we know that increased time of possession in the second half is a predictor of success. So it seems reasonable that throwing the ball more in the second half has a larger negative effect for leading teams than trailing teams. However, the fact that it has a negative effect at all for trailing teams is intriguing. The best explanation is that all teams which are behind throw the ball a lot, not just the successful teams. In the first half trailing teams call pass plays 58% of the time, in the second half this measure jumps to 65% (Table 3). It seems that perhaps the marginal effect is negative because these teams already lose 77% of the time if they are down going into the half and since they are all throwing the ball a lot it becomes negatively correlated with the probability of victory. I offer this explanation because in the first half, teams that acquire a lead obtain a significant, positive effect from increasing the percentage of pass plays. So it doesn't seem that pass plays should be all bad in the second half for the trailing team.

The final variable which had significance across both data sets was the number of times the team attempted to go for it on fourth down. For both trailing and leading teams the coefficient was negative, at around a 6% decrease in the likelihood of victory per attempt. I predicted that this statistic would be significant in the second half because it seemed more likely that teams would feel desperate and take riskier chances as the game came down to the wire. A potential explanation for the negative effect is that the only teams which go for it on fourth down need to convert, possibly late in the fourth quarter when there is little time left to score. This is obviously conjecture and needs further research. Studying where on the field and at what time teams typically go for it on fourth, as well as their success rates would make for an interesting study which would shed further light on this analysis.

B. Variables with Significance in One Set

The measure of interceptions in the first half for teams which held a lead going into the half was significant, while its counterpart for trailing teams was not. For each interception the teams probability of victory increased by 2%. The likely explanation for this is that trialing teams just typically do not get interceptions; the mean number of interceptions for leading teams is twice that of trailing teams. Alternatively, it could be that trailing teams are just not as good at converting off of interceptions and there are factors that are not accounted for in the analysis.

The number of first downs the trailing team earns in the second half significantly, negatively effects their chances of winning, while it has no effect for leading teams. The results show that every earned first down in the second half decreases the team's odds of winning by .8%. This is another somewhat surprising result. It seems to reflect the fact that first downs are actually detrimental to trailing teams. I find this dubious for similar reasons I find the negative effect of passing play percentage questionable. It could that teams really are harmed by first downs, perhaps they get a large number of first downs but are incapable of scoring. The best explanation for the negative effect may again be similar to the issue of pass play percentage, trailing teams tend to lose the game despite their best efforts and in the process of attempting to mount a comeback they typically get a lot of first downs. However, first downs may not be particularly helpful; the team needs to convert those first downs to points. This is exactly why data on red zone efficiency would be extremely helpful to my analysis, with that information it would be possible to differentiate between teams which are good at moving the ball down the field and in the process earn a lot of first downs, and teams which are good at scoring once they are in position to do so.

The percentage of pass plays was mentioned earlier, regarding the statistic in the second half. In the first half it was only significant for teams in the lead, every 1% increase led to a .22% increase in the likelihood of a win. It is somewhat difficult to explain this result. It is likely due to unaccounted for aspects of each team, since both trailing and leading teams throw a similar proportion passes in the first half and have similar completion percentages it would be expected that the percentage of pass plays thrown would have a similar effect. Perhaps another variable that could be included is a measure of the yards gained after a catch, or how far down the field the receiver travels after the reception. It could be that teams with better receivers get more value out of each throw.

Another offensive measure which only benefits the leading team in the first half is the number of passing yards. For every passing yard, the leading team increases their chance of victory by .04%. While this is an exceptionally low marginal effect it has to be taken within the context of passing. Teams with a lead at half throw for an average of 130 yards, so the effect begins to add up. I believe the explanation for the significance of this variable is similar to some of the previous ones. It is likely significant relative to the trailing team because leading teams are simply the better team most of the time and are therefore going to be better at converting yards to scoring opportunities. Additionally it is significant on its own right because a large number of passing yards means the offense is effectively moving down the field, giving the team more chances to score.

The effect of net penalty yardage is interesting because net penalty yardage is only significant in the first half for the leading team, while conversely, it is only significant in the second half for the trailing team. In the first half the leading team's chance of victory increased by .09% per positive penalty yard, in the second half the trailing team's chance of victory

increased by .18% per positive penalty yard. The best way to explain these results is by looking at the average values for each set of data. It seems clear that having positive net penalty yardage would be beneficial, since free yards help move the team downfield. On average the leading team had net positive penalty yards in the first half, while the trailing team was net negative. This trend is reversed in the second half. So it seems as though positive penalty yardage is more beneficial than negative penalty yardage is harmful, if they were equivalent we would expect to see a negative effect in the half during which the team averaged negative penalty yardage.

The final significant variable was the number of times the team attempted a two-point conversion in the second half. This measure was only significant for trailing teams and each attempt increased their odds of winning by .6%. This result likely grows out of the fact that two point conversions are rare and leading teams almost never attempt them since it is seen as an unnecessary risk. The fact that it is positively correlated with victory makes sense though; generally speaking teams go for two point conversions in a situation where doing so will win them the game so it is natural that it should increase the team's chance of victory.

C. Analysis of Marginal Effects

One, final piece of interesting analysis would be to look at comparative marginal effects of each variable on the probability of the team winning the game. This is an interesting component of the analysis because it can be difficult to conceptualize the actual effects of each separate statistic on the outcome of the game. For example, simply looking at the results in Table 4 might lead one to believe interceptions are exceptionally important, since they have the largest marginal effect per unit change. However it is important to consider some of the differences inherent in a team's production of each statistic. Practically speaking, getting an interception is

far less likely than a team gaining one extra yard from passing, so it could be that the effect of passing yards is actually typically much larger than the effect of interceptions in any given game.

To illustrate this we can consider the marginal effect of a one standard deviation change in each variable, doing so will help us understand the practical marginal effect of each variable (I will call this the standard marginal effect). This would show a type of standardized effect for each variable, which is helpful in assessing their relative importance. For teams which are behind at the end of the first half, the standard marginal effect of passing yards in the second half is 7.1%, while it is 4.6% for interceptions in the second half. Comparatively, the marginal effect of second half interceptions is 7% per interception, for passing yards it is .1%. This is exemplary of why using the standard marginal effect is useful for understanding the concrete impact of the different statistics on a team's probability of winning. The standard marginal effects for all variables are listed in Table 5.

I will discuss some of the statistics which had the largest standard marginal effects on each subset of data. As mentioned above, passing yards in the second half had a large, positive effect for both leading and trailing teams. The standard marginal effect is magnified because teams typically attain a large number of passing yards during the game. The reason why passing yards in the second half are particularly important is that they are a necessity for producing points for both trailing and leading teams.

For both types of teams, interceptions and sacks in the second half have large standard marginal effects despite their relative rarity. Though this effect is larger for the leading teams, primarily because they have larger standard deviations in the number of sacks and interceptions earned, which increases the standard marginal effect.

Finally, second half time of possession for trailing teams has a particularly large standard marginal effect. It is second only to second half passing yards. Teams which are effective at controlling time of possession can increase their odds of winning by 6% by holding onto the ball for 3.5 extra minutes in the second half. So, contrary to previous results, it seems as though time of possession is a particularly significant predictor of success. It would be interesting to run an analysis of what factors participate in significantly increasing a team's time of possession in order to further break down a team's likelihood of success in a game.

D. Insignificant Results

In general, the first half results were not as explanatorily significant as the second half. This is interesting because teams which lead at the end of the first half win far more games than teams which are trailing; which would suggest that the first half should be more significant than the second half for the use of explaining the probability that a team will win. One possible account for these results is that there is an effect of first half statistics on second half statistics. Essentially the results of the first half that would be significant in the first half are accounted for by the fact that the team is already in the lead or trailing and thus may play differently in the second half. For example completion percentage is significant in the second half but not the first for both sets of data. It is generally accepted that completion percentage is an important measure of a team's quarterbacking success, which is an important measure of the team's offensive success. It would be expected that this would be a significant measure in the first half for predicting the success of a team. However it is possible that because the teams are already divided into data sets by their first half success, this data point becomes insignificant because it is accounted for by the inherent separation of data. Similar explanations can be made for other divides between first and second half significance, or it could be that these measures are truly

important only in one half. It very well may be that time of possession only matters in the second half because of how teams strategize games.

Aside from these results, it is interesting that the number of rookies on both offense and defense is not an important factor in determining victory. This could be because not many rookies actually play in games; the average number of rookies is only 1 or 2 on each side of the ball so these individual players may not have a large impact on the game. The variable for rookie quarterbacks was close to having a negative, significant impact for leading teams; in fact it was significant at 10%. However, with the amount of data I have the effect is not particularly prominent and is only borderline significant. Though it does seem to show there could be a slight negative effect to having a rookie quarterback.

Additionally, the only variable which was not significant in either the first or second halves was the number of no-huddle attempts by a team. This may be because the no-huddle offense sees fairly limited use, with only 1-2 attempts a half regardless of the team's position. A possible fix would be to look at types of offenses and use categorical variables to distinguish between offenses which run the no-huddle often, occasionally and rarely.

VII. CONCLUSION

The goal of this study was to analyze the drivers of success in the NFL between the first and second half of games, dependent on whether the team had a lead at half. Overall, the results seems to indicate that, despite overall uniformity, there are some differences between the drivers of success for teams which hold the lead at halftime and those which do not. This could imply that NFL teams may benefit from focusing their training and resources on improving certain aspects of the game in order to improve their chances of winning.

According to the analysis of this study, key areas which teams ought to focus on are increasing passing yards, time of possession and interceptions. All of these are rather obvious and well known areas of the game which are important to victory. One area which could easily be improved by coaching and practice is the net penalty yardage. A well-disciplined team can gain a distinct advantage if they avoid committing penalties and gain positive net penalty yardage.

Additionally, teams should focus on fourth down situations. The negative effect and distribution of when in the course of a game teams attempt to convert on fourth down leads me to believe that teams which can successfully convert on fourth down would be more successful.

In the second half there are two things which particularly benefit trailing teams; time of possession and completion percentage. If a team which is behind were to focus on these two drivers alone, they would dramatically increase their chance of winning. However, breaking down statistics in the NFL is difficult because of the interconnectedness of variables. Time of possession itself is driven by many factors and a it could be an interesting topic for further study.

While the insights gained from this study are valuable and interesting, there are a multitude of further topics and ways the data could be broken down in order to provide further

and more unique insights. As mentioned multiple times, investigating how different factors affect time of possession would be one of the most interesting topics. Further, I would be interested in how interceptions are created, as there has not been much in-depth analysis of this statistic, which has a statistically large impact on the outcome of games.

VIII. REFERENCES

Arkes, Jeremy. "Is Controlling the Rushing or Passing Game the Key to NFL Victories?" Contemporary Sports Issues, Sports Coaching, Sports Management, Sports Studies and Sports Psychology, February 2, 2011. Accessed March 30, 2016. http://thesportjournal.org/article/is-controlling-the-rushing-or-passing-game-the-key-to-nfl-victories/.

Atkinson, Scott E., Linda R. Stanley, and John Tschirhart. "Revenue Sharing as an Incentive in an Agency Problem: An Example from the National Football League." The RAND Journal of Economics 19, no. 1 (1988): 27-43. Accessed March 30, 2016. doi:10.2307/2555395.

Burton, William, and Michael Dickey. "Statistical Model Predicts with High Accuracy Play-calling Tendency of NFL Teams." EurekAlert! Accessed April 24, 2016. http://www.eurekalert.org/pub_releases/2015-08/asa-smp080615.php.

Hadley, Lawrence, Marc Poitras, John Ruggiero, and Scott Knowles. "Performance Evaluation of National Football League Teams." Managerial and Decision Economics Manage. Decis. Econ. 21, no. 2 (March 2000): 63-70. Accessed March 30, 2016. doi:10.1002/1099-1468(200003)21:23.0.co;2-o.

Hofler, Richard A., and James E. Payne. "How Close to Their Offensive Potential Do National Football League Teams Play?" Applied Economics Letters 3, no. 11 (1996): 743-47. Accessed March 30, 2016. doi:10.1080/135048596355808.

Hu, Bo, Mari Palta, and Jun Shao. "Properties of R2 Statistics for Logistic Regression." Statistics In Medicine 25, no. 8, 1383-395. Accessed April 14, 2016. doi:10.1002/sim.2300.

Isidore, Chris. "NFL Revenue: Here Comes Another Record Season." CNNMoney. Accessed March 30, 2016. http://money.cnn.com/2015/09/10/news/companies/nfl-revenue-profits/.

Lewis, Michael. *Moneyball: The Art of Winning an Unfair Game*. New York: W.W. Norton, 2003.

McGee, Kevin, Lee Van Syoc, and Nancy J. Burnett. "'The Best Defence…' or Optimal Offence/defence Spending Ratios in the NFL." Applied Economic Letters 19, no. 8 (September 2012): 717-20. Accessed March 30, 2016. doi:10.1080/13504851.2011.597715.

Pallotta, Frank, and Brian Stetler. "Super Bowl 50 Audience Is Third Largest in TV History." CNNMoney. Accessed March 30, 2016. http://money.cnn.com/2016/02/08/media/super-bowl-50-ratings/.

Pitts, Joshua D. "Determinants of Success in the National Football League's Postseason: How Important Is Previous Playoff Experience?" Journal of Sports Economics 17, no. 1 (January 2016): 86-111. Accessed March 30, 2016. doi:10.1177/1527002514525409.

Robst, J., VanGilder, J., Berri, D. J., & Vance, C. (2011). 'Defense wins championships?': The answer from the gridiron. International Journal of Sport Finance, 6(1), 72-84. Retrieved from http://ccl.idm.oclc.org/login?url=http://search.proquest.com/docview/863850682?accountid=10141

Stern, Hal. "On the Probability of Winning a Football Game." The American Statistician 45, no. 3 (1991): 179. Accessed March 30, 2016. doi:10.2307/2684286.

XI. TABLES

Table 1.
List of Variables with Descriptions

Variable Name	Description
home	Binary variable, 1 if the team is playing at home, 0 if the team is away
streak	The number of games the team has won or lost in a row, resets to 0 after a streak is broken
qbrookie	Binary variable,1 if the quarterback is a rookie, 0 if otherwise
offrookies	The number of offensive rookies on the team who participate in the game
defrookies	The number of defensive rookies on the team who participate in the game
sackfor1	The number of sacks earned by the team's defense in the first half.
sackfor2	The number of sacks earned by the team's defense in the second half.
intsfor1	The number of interceptions earned by the team's defense in the first half.
intsfor2	The number of interceptions earned by the team's defense in the second half.
fumrec1	The number of by the team's defense in the first half.
fumrec2	The number of fumbles recovered earned by the team's defense in the second half.
startfieldpos1	The average starting field position of the team in the first half. This is calculated by taking the starting field position at the beginning of each drive, adding it up then dividing by the number of drives.
startfieldpos2	The average starting field position of the team in the second half.
fd1	The number of first downs earned by the team in the first half.
fd2	The number of first downs earned by the team in the second half.
percentpassplay1	The percentage of plays by the team which were passing plays, in the first half. This was calculated by dividing the total number of pass plays in a half by the total number of running and passing plays in the half. Since this is a percent, the coefficient in the regression is already in percentage amount.
percentpassplay2	The percentage of plays by the team which were passing plays, in the second half.
ry1	The number of rushing yards by the team in the first half.
ry2	The number of rushing yards by the team in the second half.
py1	The number of passing yards by the team in the first half.
py2	The number of passing yards by the team in the second half.

comp1	The percentage of passes completed by the team in the first half. Since this is a percent, the coefficient in the regression is already in percentage amount.
comp2	The percentage of passes completed by the team in the second half.
nh1	The number of no-huddle plays called by the team in the first half.
nh2	The number of no-huddle plays called by the team in the second half.
top1	The time of possession (in minutes) by the team in the first half.
top2	The time of possession (in minutes) by the team in the second half.
netpenydg1	The net penalty yardage for the team in the first half. This was calculated by subtracting penalty yards against the team from penalty yards for the team.
netpenydg2	The net penalty yardage for the team in the first half.
go4th1	The number of times the team attempts to convert to first down while on fourth down, in the first half.
go4th2	The number of times the team attempts to convert to first down while on fourth down, in the second half.
conv1	The number of times the team attempts a two point conversion in the first half.
conv2	The number of times the team attempts a two point conversion in the second half.

Table 2.
Summary Statistics for Teams Leading at Halftime

Variable	Obs	Mean	Std. Dev.	Min	Max
home	3,909	.562548	.4961358	0	1
streak	3,909	.3289844	2.856735	-13	18
qbrookie	3,909	.0918393	.2932321	0	2
offrookies	3,909	1.398056	1.087124	0	7
defrookies	3,909	1.882067	1.302764	0	8
sackfor1	3,909	1.25275	1.143076	0	9
sackfor2	3,909	1.445382	1.293792	0	8
intsfor1	3,909	.6068048	.7463077	0	4
intsfor2	3,909	.7086211	.8356684	0	5
fumrec1	3,909	.5126631	.6940718	0	4
fumrec2	3,909	.4704528	.6935581	0	5
startfield~1	3,909	31.07136	7.913918	7	60.28571
startfield~2	3,909	31.655	8.454545	7.25	58.16667
fd1	3,909	11.54899	3.371181	1	26
fd2	3,909	9.362241	3.106645	1	24
percentpas~1	3,909	.5501448	.1043517	.125	.9310345
percentpas~2	3,909	.4689222	.1557319	.03125	.9285714
ry1	3,909	66.66283	33.23844	-6	251
ry2	3,909	63.81069	35.47845	-10	233
py1	3,909	129.4612	53.47101	-3	357
py2	3,909	90.94372	47.4529	-18	306
comp1	3,908	1.720031	.4883967	1	10
comp2	3,909	.5758089	.1537543	0	1
nh1	3,909	1.59427	4.475068	0	39
nh2	3,909	1.087746	3.514442	0	33
top1	3,909	16.50544	2.936277	6.533333	26.6
top2	3,909	16.20324	3.279173	4.133333	32.51667
netpenydg1	3,909	.6999233	16.16589	-50	82
netpenydg2	3,909	-1.225889	15.48071	-69	106
go4th1	3,909	.146329	.3799948	0	3
go4th2	3,909	.231517	.5112544	0	4
conv1	3,909	.0115119	.1158851	0	3
conv2	3,909	.0608851	.247562	0	2

Table 3.
Summary Statistics for Teams Trailing at Halftime

Variable	Obs	Mean	Std. Dev.	Min	Max
home	3,909	.437452	.4961358	0	1
streak	3,909	-.1097467	2.744116	-15	15
qbrookie	3,909	.1343054	.3499121	0	2
offrookies	3,909	1.561269	1.155577	0	7
defrookies	3,909	1.961883	1.325998	0	9
sackfor1	3,909	.8963929	.9786669	0	6
sackfor2	3,909	.9035559	1.023351	0	7
intsfor1	3,909	.2921463	.5243236	0	3
intsfor2	3,909	.3747762	.6442313	0	4
fumrec1	3,909	.2778204	.5286778	0	4
fumrec2	3,909	.3632643	.5998475	0	4
startfield~1	3,909	26.89465	6.455625	8.666667	54.14286
startfield~2	3,909	28.82453	7.690465	8.25	62.33333
fd1	3,909	8.160399	3.045561	0	19
fd2	3,909	10.56792	3.73506	0	29
percentpas~1	3,909	.5836412	.1029559	.0714286	.9285714
percentpas~2	3,909	.6519337	.1454354	.09375	1
ry1	3,909	47.99923	26.57454	-12	197
ry2	3,909	48.18419	31.94641	-14	311
py1	3,909	88.85162	45.24252	-19	290
py2	3,909	124.7229	59.01957	-22	361
comp1	3,908	2.040584	.6997666	1	11
comp2	3,909	.5484788	.1278701	0	1
nh1	3,909	1.188539	3.454709	0	32
nh2	3,909	2.214889	4.60451	0	36
top1	3,909	14.36256	2.945447	4.016667	24.13333
top2	3,909	14.98825	3.513819	4.566667	30.98333
netpenydg1	3,909	-1.876951	15.10771	-60	79
netpenydg2	3,909	.4177539	16.86722	-75	93
go4th1	3,909	.1412126	.3772071	0	3
go4th2	3,909	.7234587	.8767524	0	5
conv1	3,909	.0079304	.0915495	0	2
conv2	3,909	.1916091	.4623725	0	5

Table 4.
Marginal Effects for Probit Models with Winning as Dependent Variable

VARIABLES	(1) lead@half	(2) behind@half
home	0.0514*** (0.0112)	0.0502*** (0.0106)
streak	0.00669*** (0.00186)	0.00707*** (0.00185)
qbrookie	-0.0325* (0.0185)	-0.0180 (0.0169)
offrookies	-0.00400 (0.00509)	-0.00497 (0.00457)
defrookies	0.00407 (0.00401)	0.00105 (0.00375)
sackfor1	-0.00166 (0.00475)	0.00101 (0.00489)
sackfor2	0.0374*** (0.00491)	0.0358*** (0.00485)
intsfor1	0.0204*** (0.00784)	-0.00170 (0.00927)
intsfor2	0.0681*** (0.00816)	0.0717*** (0.00832)
fumrec1	0.00509 (0.00853)	0.00710 (0.00982)
fumrec2	0.0517*** (0.00963)	0.0479*** (0.00814)
startfieldpos1	0.00134* (0.000766)	0.00117 (0.000801)
startfieldpos2	0.00345*** (0.000760)	0.00310*** (0.000727)

fd1	0.00316 (0.00299)	-0.00171 (0.00312)
fd2	-0.000131 (0.00301)	-0.00762*** (0.00259)
percentpassplay1	0.229*** (0.0709)	0.0290 (0.0661)
percentpassplay2	-1.175*** (0.0767)	-0.806*** (0.0683)
ry1	0.000222 (0.000246)	0.000165 (0.000280)
ry2	-0.000299 (0.000246)	0.000398* (0.000233)
py1	0.000455*** (0.000164)	0.000131 (0.000177)
py2	0.00108*** (0.000188)	0.00121*** (0.000151)
comp1	0.000810 (0.0115)	-0.0107 (0.00951)
comp2	0.161*** (0.0466)	0.330*** (0.0523)
nh1	0.00143 (0.00211)	0.00238 (0.00205)
nh2	-0.00203 (0.00238)	-0.00141 (0.00163)
top1	-0.00339 (0.00238)	0.00145 (0.00229)
top2	0.00758*** (0.00208)	0.0170*** (0.00187)
netpenydg1	0.000889** (0.000357)	0.000555 (0.000350)

netpenydg2	0.000430	0.00176***
	(0.000355)	(0.000310)
go4th1	-0.00218	-0.0130
	(0.0145)	(0.0137)
go4th2	-0.0561***	-0.0686***
	(0.0108)	(0.00764)
conv1	-0.0110	0.0567
	(0.0406)	(0.0493)
conv2	0.00686	0.0290***
	(0.0193)	(0.00968)
McFadden Pseudo R^2	.5436	.5494
Observations	3,908	3,908

Table 5.

Standard Marginal Effects of Variables (standard deviation times the marginal effect)

Lead at Halftime			Behind at Haltime		
Variable	**std.*mfx**	**Singificant (α=.05)**		**std.*mfx**	**Singificant (α=.05)**
comp1	0.0%	NO	comp1	-0.8%	NO
comp2	2.5%	YES	comp2	4.2%	YES
conv1	-0.1%	NO	conv1	0.5%	NO
conv2	0.2%	NO	conv2	1.3%	YES
defrookies	0.5%	NO	defrookies	0.1%	NO
fd1	1.1%	NO	fd1	-0.5%	NO
fd2	0.0%	NO	fd2	-2.8%	YES
fumrec1	0.4%	NO	fumrec1	0.4%	NO
fumrec2	3.6%	YES	fumrec2	2.9%	YES
go4th1	-0.1%	NO	go4th1	-0.5%	NO
go4th2	-2.9%	YES	go4th2	-6.0%	YES
home	2.5%	YES	home	2.5%	YES
intsfor1	1.5%	YES	intsfor1	-0.1%	NO
intsfor2	5.7%	YES	intsfor2	4.6%	YES
netpenydg1	1.4%	YES	netpenydg1	0.8%	NO
netpenydg2	0.7%	NO	netpenydg2	3.0%	YES
nh1	0.6%	NO	nh1	0.8%	NO
nh2	-0.7%	NO	nh2	-0.6%	NO
offrookies	-0.4%	NO	offrookies	-0.6%	NO
percentpassplay1	2.4%	YES	percentpassplay1	0.3%	NO
percentpassplay2	-1.8%	YES	percentpassplay2	-1.2%	YES
py1	2.4%	YES	py1	0.6%	NO
py2	5.1%	YES	py2	7.1%	YES
qbrookie	-1.0%	NO	qbrookie	-0.6%	NO
ry1	0.7%	NO	ry1	0.4%	NO
ry2	-1.1%	NO	ry2	1.3%	NO
sackfor1	-0.2%	NO	sackfor1	0.1%	NO
sackfor2	4.8%	YES	sackfor2	3.7%	YES
startfieldpos1	1.1%	NO	startfieldpos1	0.8%	NO
startfieldpos2	2.9%	YES	startfieldpos2	2.4%	YES
streak	1.9%	YES	streak	1.9%	YES
top1	-1.0%	NO	top1	0.4%	NO
top2	2.5%	YES	top2	6.0%	YES

Table 6.
4th Down Conversion Attempts by Team Type and Half

Team Type	1st half	2nd half
Leading	12%	19%
Trailing	11%	58%
Total	23%	77%