

The Effect of Luck on the Retention of Head Coaches in the NFL

A Thesis By:

Alex Steen

April 5th, 2016

In Partial Fulfillment of the Requirements for the Award of M.S. in Sport Management
and Administration

Copyright © 2016
Alexander Craig Steen
All Rights Reserved

Thesis Committee

Dr. Kevin Heisey

5 April

Dr. Jeffrey Briggs

5 April

Dr. Phil Blosser

5 April

Abstract

This investigation considers the impact of the concept of luck on coaching retention in the National Football League (NFL). The head coach of an NFL team is often under intense scrutiny to perform in a high profile position. Due to a small inventory of games each season, there is the potential for luck to influence whether or not a season is viewed a success. This study seeks to measure the impact of luck for a given NFL team while attempting to correlate with the retention or release of the head coach.

Keywords: retention, coaching, NFL, luck, variance, sport management

TABLE OF CONTENTS

CHAPTER ONE: INTRODUCTION.....	1
1.2 Purpose of the study.....	1
1.2.1 Research Objectives.....	2
1.3 Defining Luck.....	2
CHAPTER TWO: LITERATURE REVIEW	4
2.1 Coaching.....	4
2.2 Pythagorean Win Expectation.....	11
2.3 Why the NFL?.....	14
2.4 Luck	17
CHAPTER THREE: METHODOLOGY	24
3.1 Building the model.....	25
3.2 The data set	33
CHAPTER FOUR: RESULTS.....	36
4.1 Data set analysis.....	36
4.2 Regression of the model.....	38
4.3 Regression to isolate bad luck.....	41
4.4 Regression to isolate good luck	42
CHAPTER FIVE: DISCUSSION.....	45
5.1 Data set analysis discussion	46
5.2 Initial regression discussion.....	48
5.3 Discussion of luck regressions.....	50
5.4 Further Research	51
5.5 Thinking outside the norm	53
5.6 Conclusion	53

LIST OF TABLES

Table 1. Model Variables	31
Table 2. Dependent Variable Frequency	36
Table 3. Descriptive Statistics - Variables	36
Table 4. Descriptive Statistics - Cases	37
Table 5. Regression Results - Model.....	39
Table 6. Dependent Variable Frequency	39
Table 7. Overall Descriptive Statistics.....	39
Table 8. Regression Results – isolated for bad luck	41
Table 9. Dependent Variable Frequency	41
Table 10. Bad Luck Regression Descriptive Statistics	42
Table 11. Regression Results – isolated for good luck.....	43
Table 12. Dependent Variable Frequency	43
Table 13. Good Luck Regression Descriptive Statistics.....	43
Table 14. Good and bad luck teams.....	50

CHAPTER ONE: INTRODUCTION

Sport is often considered big business at the highest levels (Frederick, 2013). Similar to the management of any other high profile business, sport management is highly visible and quite public often leading to public debate and scrutiny of management decisions. In a high visibility environment, it can be difficult to deviate from accepted logic. Going against commonly held beliefs could result in a loss of support from fans and sponsors. Winning may gain back some of that support, but if luck is a significant contributing factor the right decision may still yield negative results. The potential for the correct decision to yield negative results makes any thinking outside of the norm both a potential break through and a potential disaster for a sport management team.

One of the most high profile decisions a management team in the National Football League (NFL) must make is whether or not to fire a head coach. There are any number of factors, both quantitative and qualitative, to consider when making such a decision. A coach's job performance could be measured in many different ways across many different areas. Some of these areas may include whether the team is performing well on the field (Holmes, 2010), whether the fans and sponsors support the coach, or whether the management team believes the coach has the ability to do the job (Mason, 2014). All of these factors may be considered when determining coach retention. One factor that may be overlooked by some is the concept of luck.

1.2 Purpose of the study

The purpose of this study is to investigate the luck factor and its impact on the decision to retain NFL head coaches. With the rise of advanced analytics, there is likely

more information for a management team to consider now than ever before when making such a decision. While this could make determining how a coach is performing easier, if there is a significant factor in a coach's performance that is not considered, management teams could be operating under less than ideal conditions. It is the aim of this study to determine if the impact of luck on an NFL team could determine the employment fate of the head coaches.

1.2.1 Research Objectives

- 1) To gain an understanding of the impact of luck on an NFL team's outcomes during a given season.
- 2) To determine how much, if any, effect the impact of luck has on NFL head coach's retention.

1.3 Defining Luck

In order to determine if luck is a significant contributing factor, it is necessary to assign it an operational definition. Michael Mauboussin, Chief Investment Strategist at Legg Mason Capital Management, describes luck possessing having three features:

- 1) The first is that it occurs to both groups and individuals.
- 2) Second, the effect can be good or bad.
- 3) Third, luck may play a role when it is reasonable to believe that something else may have happened (Arbesman, 2012).

If you've watched sport on television you likely have heard it said of a team that just lost, "If they played that game ten more times, they would win nine" (Bishop, 2009). While that colloquialism is hardly scientific, there are plenty of examples of one team statistically defeating another team, while losing the contest. If these aberrant results happen to cluster together, a team's win-loss record for a stretch of games may not reflect the true performance of the team. This sort of variation tends to even out over time. However, until it does, it has the effect of either good or bad luck, depending on which way the variation goes. If a management team happens to make a decision based on data sets with small sample sizes and the data set has been skewed due to this luck factor, the decision could cause people to lose their jobs erroneously. This suggests that, for the purposes of this study, luck and variation are closely intertwined. Thus, they will be treated the same.

CHAPTER TWO: LITERATURE REVIEW

2.1 Coaching

One of the most visible decisions a management team has to make is the decision to retain or fire a head coach (Frederick, 2013). In this decision making process the performance of the coach must be weighed against the likely performance of the replacement. While it is easy for a team that is not winning to consider making a change at the top, some research suggests changing coaches results in worse performance than keeping the incumbent. This is likely due to the turnover in personnel and change in strategy accompanying a new coach (Adler, Berry & Doherty, 2013). While a change in approach may be attractive to a management team when their team is not winning, it also tends to interrupt the development of an organization. Deciding on whether or not a change of approach is necessary is complicated, and therefore, the team's performance alone may not be an adequate metric for coach dismissal.

While a team's performance alone may not be enough to dismiss a coach, every situation is different (Adler, Berry & Doherty, 2012). Even in professional football, where there is a salary cap and a competitive framework encouraging parity, each management team will value some aspects of performance over others (Deceuster & Kishor, 2012). This complex situation is difficult enough to navigate without any luck at all involved. If luck is a significant factor, it adds another layer of complexity. Despite the complexity of probability factors and luck, there are simple ways to evaluate it. These ways may over-simplify a situation, but can be a starting point to launch a broader

discussion. Rating a coach's performance could be as simple as measuring wins and losses against what is expected by the organization (Adler, Berry & Doherty, 2012). While some teams may only be expected to win more than lose, others may be expected to win championships. A coach's true performance level cannot be entirely measured by winning percentage. Before the rise of advanced metrics, management teams had to make judgment calls about player talent and the effect of injuries. Based on their opinions of those and other metrics, they could then judge the performance of the coach. In more recent times, researchers such as Brian Burke have used win probability formulas to estimate how a coach's decision making effected the number of wins and losses his team recorded (Meiselman, 2012). These measures vary widely in results and what they attempt to measure. This suggests that a complete and accurate predictor of good coaching has not yet been found. One of the major difficulties in measuring predictors of successful coaching is accurately gauging player talent (Dawson, Dobson & Gerrard, 2000). Certain players will thrive in certain systems while floundering in others making it difficult to determine whether a team's struggles are due to a coaching deficiency or some other variables.

The question of player quality is another layer of complexity that must be addressed. The acquisition and retention of players also brings the politics of a management team into play. In the NFL, the control over personnel decisions is often shared by many different people and varies from team to team (Breer, 2013). When a head coach has little involvement in personnel decisions, it would be reasonable to believe the persons involved in organizational personnel may share some of the blame when a team fails. While that may well be the case, that person has a vested interest in

directing blame towards a coaching staff, and away from the players, as his own job may be at stake if he does not. This power struggle could have an impact on the decision making process, which is difficult to quantify. What is clear is that each situation is different and understanding who holds the power over what is important in evaluating performance (Breer, 2013). Taking an analytical approach and looking purely at the hard numbers allows us to discern which variables have the greatest impact, but we should not ignore that there is always more to the story that numbers have a hard time capturing (Williams, 2015).

There is much more literature available on the effect of coaching changes at the collegiate level than there is on the effect in the NFL. Maxcy (2013) suggested the sport of football may differ from the accepted logic of the marketplace considering favoring retention. In that study it was found that the replacement of coaches whose teams were not performing was usually a good decision. With college football having far less parity than any professional sport and the NFL having more parity than the other major professional leagues it would be reasonable to expect changes to have a greater impact in the NFL (Deceuster & Kishor, 2012). Add to that the existence of a salary cap in the NFL ensures that the difference in talent from team to team is much less than in college football and the importance of a good coach becomes evident. Therefore, if the trend holds from college football to the professional ranks, it would be even more important for NFL teams to make a change when their teams aren't performing.

The above discussion suggests why the effect of coaching in the NFL may be magnified compared to lower levels of the game. If that is the case, finding a measure of how effective a particular coach is would be a valuable tool. Hadley et al. (2000),

developed such a measure and found the mean efficiency rate of NFL head coaches to be .641. This is a coaching efficiency percentage. Thus, if a perfect coach would be able to produce 10 wins over the course of a 16 game NFL season with a given roster, the average head coach would only be able to produce 6.41 wins in the same situation. While a perfect coach likely does not exist, this is a dramatic difference considering 10 wins usually qualifies a team for the play-offs while 6 or 7 wins is a losing season. When you also consider that there are below average coaches out there you find that some teams may have no chance at a successful season regardless of talent. The Hadley study looked at statistics that encompass on field performance in each phase of the game and used those to predict the number of wins a team should have had.

While some might suggest that head coaches also have a significant impact on player statistics, such as those used in the Hadley study to predict the number of wins a team should have had, other research does not find that to be the case. When using somewhat simplistic measures to measure coaching ability, years of experience or win-loss record, coaching does not appear to have a significant impact on player's statistics (Berri & Schmidt, 2006). This finding, in addition to the rather impressive list of names that Hadley found to have the highest coaching efficiency ratings, suggests that head coaches have a great deal of impact on the number of wins a team achieves, even if their impact on individual player's statistics is negligible. That one coach could achieve drastically different results given the same level of player talent would lend credence to those who call for the head coaches firing when their team isn't winning games.

Despite the suggestion that coaches have a great deal of impact on the performance of their team, there are many factors that remain outside a coaches' control.

While the head coach has long been held accountable for the performance of his team, his players have a great deal of impact on that performance. Kahn (2006) found that the effectiveness of a coach is based on the quality of his players. In the NFL, it is rare for a head coach to have complete control over the players on his team (Breer 2013). When that is the case, it may impact the evaluation of the coach's performance. Management could evaluate the quality of players a coach has to work with and take that into account when evaluating whether a coach is performing adequately.

There needs to be further examination of the relationship between good coaches and player performance. Proving how much of a team's on-field success is due to the players and how much is due to the coaches is, obviously, an extremely difficult task, but doing so would give management teams a huge advantage in making personnel decisions (Dawson & Gerrard, 2000). Knowing when a team has the right coach, but the wrong players, or the wrong coach, but the right players, would eliminate the potential to replace the wrong people. Replacing the wrong people, in that situation, could be devastating to a franchise. Making the right call should be the focus of every management team looking to improve their squad. Until data analysis researchers can figure this problem out, management teams will still have to rely on their knowledge and experience to, hopefully, get it right.

There are a number of examples of coaches that failed in one situation only to go on and succeed in another. Bill Belichick had a career record below .500 in the NFL when the Patriots hired him in 2000 ("Bill Belichick", 2015, p. 1). He has gone on to be one of the most successful coaches in NFL history. The quality of players may be to blame, but there is another possibility. Porter and Scully (1982) suggest that baseball

managers are more successful as they gain experience. This learning curve based theory makes sense when applied to any job where experience factors into performance. If management teams understand this concept, then they will inherently expect less from a new head coach. This would obviously apply to a new head coach with no prior experience, but it should also apply to any new head coach taking over a team with which he is unfamiliar (Porter & Scully, 1982). Because this involves learning about his new players, becoming familiar with a new situation, and may involve building a new coaching staff, a learning curve exists regardless of prior experience levels. In football, this could be even more pronounced than in baseball. New head coaches may implement an entirely different offensive and defensive scheme which complicates getting to know new players and may necessitate a higher rate of personnel turnover than usual.

This grace period early in a coach's tenure, to both learn and shape his team, should not be extrapolated to mean that a coach will continue to improve throughout his tenure. While he will continue to gain experience, the law of diminishing returns could apply here (Porter & Scully, 1982). In addition, once his personnel are in place and he has a team playing as he likes, the coaches' upward trend in performance should flatten substantially. This effect may be compounded if, as is the case with many successful teams, key assistant coaches are hired away by other teams. Add to that the potential for becoming stale, both with players and approach to the game, and coaching performance could even regress. The argument to allow a coach time to grow should not be interpreted as an argument to retain a coach indefinitely.

The idea that new head coaches need to be given time to adapt can also be reversed. Management teams also need time to adjust to the new head coach and to

evaluate his performance (Scully, 1994; Kahn, 2006). The management team that hired the new coach will inherently believe that he can be successful so it makes sense that they would give him time to develop while they build their opinions of his performance. As head coaches turnover, so do management teams. A change in management teams may negate this grace period or extend it depending on the new team's attitude towards the coach.

All of these things could apply to any franchise that hires a new head coach. In addition, the same line of thinking may help to explain why some coaches keep getting opportunities even after several failed tenures with various teams. Because the evaluation of a coach's performance is such a difficult task, there is almost always an argument that prior failures had more to do with factors beyond his control than his own problems (Mihoces, 2015). That, in addition to the possibility that coaches may improve with experience as discussed in regard to the learning curve theory, may lead teams to give failed coaches second, third, or even fourth chances. If luck is a significant factor in some coaches getting fired, that would bolster the argument that these failed coaches could still be worth the risk.

The effects of a coach on a team's performance, the effects of a coaching change, and how to evaluate a coaches' performance have all been well studied, both in the academic world and in less rigorous environments. Such extensive study of this particular decision suggests that many people see it as an important one. In addition, this research has established that the decision whether to fire a coach or not is a complex one. With so many variables to consider, many of them difficult to quantify, the possibility that luck is a significant factor in ultimately determining whether a coach keeps his job must

be considered.

2.2 Pythagorean Win Expectation

In 1977 Bill James published *Baseball Abstract*. In that publication, James introduced several statistics that he had designed himself to further data analysis focused on baseball. One of the statistics he introduced was the Pythagorean Win Expectation (PWE) formula (Luo, 2014, p. 5). The PWE is a relatively simple formula. Its only inputs are the number of points scored and the number of points allowed. Though it was introduced to analyze baseball, it has since been adapted to the NFL and other major sports.

The purpose of the PWE is to predict a team's winning percentage for a given season based on their points scored and points allowed. There has been further research into improving it, but the original PWE formula has proven to be nearly as accurate as the best others have come up with (Luo, 2014, p. 1). In that research, there were several different attempts made to use advanced metrics to improve the PWE formula in an attempt to improve its predictive accuracy. The PWE average prediction for a team's wins was 4.22 games off for a 162 game baseball season. The best of the improved formulas averaged 3.11 games off (Luo, 2014, p. 5). This difference was found to be statistically significant. However, the difference, about 25%, is not of a magnitude that it would negate the usefulness of the PWE when a simpler formula is called for. This supports both the idea that the PWE is not a perfect formula and also the idea that it is accurate enough to use for this sort of study.

To that point, one potential issue with the PWE formula is that it tends not to predict teams finishing with records to either extreme (Luo, 2014). Because of the way the expected winning percentage is calculated, it is difficult, if not impossible, for the

predictions to be correct for teams that finish with very few wins or losses respectively. For example, the 2007 New England Patriots set an NFL record for the highest point differential in one season in the Super Bowl Era. They went undefeated in the regular season, outscoring their opponents 589-274 (“2007 New England Patriots”, n.d., p. 1). However, those Patriots would have only been expected to win 13.76 games using the PWE. To illustrate the other side of the coin, when the Detroit Lions lost every game of the 2008 season, they were outscored 268-517 (“2008 Detroit Lions”, n.d., p. 1). The PWE would have projected them to have won 2.78 games that season. As you can see, teams with very good or very bad records seem to cause the PWE some trouble.

The purpose of using the PWE for this study is to provide a baseline. The PWE predicts how many games a team won, over a given stretch, based upon their on-field performance (Luo, 2014). This number will be used as a baseline for how a team performed in a given season. Once a baseline has been established, the difference between that baseline and the actual number of wins a team earned can be examined. I would expect that most teams would end up fairly close to the baseline with their final win total. That is the idea behind the PWE and it has been shown to be accurate enough to support that expectation (Luo, 2014). Teams that end up fairly close to the PWE predicted baseline are not interesting in this context. They represent teams that fit into the statistical model well and have average luck. Of course, they will still be included in the data set, but drawing conclusions about luck from them is difficult if not impossible.

Instead of those teams that fall close to their expected win total, it is more interesting to examine those teams that experienced results that fall towards the extreme bounds of what we might expect due to variation. While these could be outliers, and due

to something other than luck, research discussed later indicates a strong tendency for these teams to regress to the mean which supports the idea that this is likely variation and luck at work (Barnwell, 2013). Teams that either far exceed or fall far short of the win total predicted by the PWE may represent those who either have exceptional luck, good or bad, or those that do not fit the statistical model well for one reason or another (Barnwell, 2013). I am interested here in the luck aspect. Later, I will discuss some of the evidence that these outliers may have luck to blame, but for now the possibility that it could be luck and the role that the PWE plays in determining that is enough.

This is not the first time that the PWE has been used to judge how lucky a team was over the course of a season. A simple internet search turns up analysis of how luck has effected Nebraska football over the years (Vogel, 2012) and which Big 10 football teams had the best luck (Vint, 2011). These articles are, generally, basic analysis presented for average fans. However, with PWE being consistently used as a way to tell how lucky a team was, it begs the question as to why. That is an easy question to answer if the PWE accurately predicts what a team's winning percentage should have been over a number of games. If the PWE gives us a decent idea of what a team's win-loss record might have been, we can measure the difference between that and their actual win-loss record to determine if the sample was skewed towards more or less wins. Put another way, whether they had good or bad luck.

While that is a relatively simple determination, it relies on an under-lying assumption. That the PWE is able to accurately give us that baseline we need to measure from. Caro and Machtmes (2013) investigated just that. Building on other research that had investigated what the optimal exponent for the PWE is for each sport, they

investigated how the PWE fit college football results in an effort to determine if the PWE is applicable to college football and if it is useful as a retrospective tool. They even posit, much as I do, that “it could prove to be an effective tool for athletic directors to use when making coaching decisions” (Caro & Machtmes, 2013, p. 538). Ultimately, Caro and Machtmes found that the PWE is “an accurate forecasting method for coaches and athletic directors to use when evaluating the performance of their football programs” (Caro & Machtmes, 2013, p. 538).

Caro and Machtmes go on to opine about how a coach might use the PWE and the difference between it and actual winning percentage to draw conclusions about his team and his own decision making. This is certainly an area that needs further study, but Bill Barnwell's work, which will be discussed later, suggests that simply waiting for the sample size to grow might be enough for bad luck to even out (Barnwell, 2013). This could be one of the most useful aspects of the PWE. If a team can identify when they have been unlucky, they may be able to avoid overreacting and changing their approach when it is unnecessary. This is why I use the PWE in this study. I want to find out if coaches are being fired when the PWE suggests staying the course might result in performance that a management team would find acceptable.

2.3 Why the NFL?

That the PWE was initially a baseball measure is not surprising. Baseball was the first sport to see the rise of advanced statistics. Because of its long season and the make-up of the game, baseball is the most quantifiable sport (Moy, 2006, p. 2). For most studies, this is an advantage. Every pitch in a baseball game can be recorded. While new research into defensive metrics, such as range factor, have helped cast new light on other metrics,

in general, there can be more useful analysis done on baseball without actually watching the games than a sport like football (Moy, 2006). Football may be more difficult to analyze due to its many moving parts. While each play in baseball begins as a battle between the pitcher and batter, each play in football begins with 11 players on offense battling 11 players on defense. Trying to identify the goals of each player on any given play is difficult enough. Trying to analyze how their success or failure impacts the other battles on the field and, ultimately, the result of the play can be quite difficult. All of this suggests that baseball is a better venue for research than football (Moy, 2006, p. 2). However, this study takes a higher level view of on-field performance. Because I am only interested in a team's winning percentage, their points scored, and their points allowed, the play by play details are much less important.

If the play by play details of the sport are unimportant, what is important?

Because I am interested in the effect luck may have on a coach being fired, I need to find a sport where luck has the maximum impact on a season's result. To find such a sport we must examine what aspects of a data set increase variance. Variance is a measure of the contribution to error of deviations from the central tendency (Brain & Webb, 1999, p. 2). I am looking for a sport with high variance because, in the short term, a sport with higher variance is more influenced by luck. Over the long term, luck will even out and a team's results will naturally converge on where they should be based on the team's performance (Barnwell, 2013). In this context, short term and long term refer more to the sample size than any sort of time measurement. Thus, we are looking for a sport that has the smallest sample size of outcomes.

The NFL, more so than any other major professional sport in the US, uses a very

small sample size of outcomes to determine a team's standing in any given year. If a team over-performs by just 2 wins or under-performs by the same, that team could be seen much differently than their overall performance would suggest. Preliminary investigations into such teams suggest that a large differential between wins and wins predicted by the PWE usually predicts that a team will regress towards the PWE predicted level the next season (Barnwell, 2013, p. 1).

Barnwell's work is a good example of high level football analytics work being done on commercial websites, previously on Football Outsiders and now on ESPN.com's Grantland. Commercial websites present an interesting challenge when operating in an academic environment. The level of research being done by these websites is respected by the NFL teams themselves which suggests that it should not be dismissed simply because it doesn't come from academia (Vrentas, 2015, p.1). While the work on commercial sites fails certain tests, it isn't typically peer-reviewed before being published for instance, the competition that now exists between sites and analysts to provide analytical services to NFL teams themselves suggests that these sources be taken seriously. Pro Football Focus is now providing data to 13 NFL teams (Vrentas, 2015, p. 1). The Denver Broncos hired a former STATS LLC employee to lead their work in this area as a director of football analytics (Swanson, 2015, p. 1). This community of high level football analysis feeds off of each other's work and provides a sort of informal peer review. Bill Barnwell and Brian Burke, in particular, often refer to each other's work to support their own ideas (Barnwell, 2013). Between the sites already mentioned, Brian Burke's site, and Advanced Football Analytics, the work being done is as good as any and can fill many of the gaps left by traditional academic research in this field.

Going back to Barnwell's finding that a team's win total will typically regress towards their PWE predicted total, it fits with the suggestion that the difference between actual wins and the win total expected by the PWE is, at least in part, due to luck (Barnwell, 2013). As previously noted, a larger sample size would see outliers regress towards the type of results their performance deserved. This appears to be exactly what is happening to teams that have large differentials. This suggests that luck is a factor over an NFL season. If that is the case, that luck could, potentially, be the difference between a coach getting fired or retained. If a decision is made before that luck balances out, the wrong decision could be made.

2.4 Luck

In this study I discuss luck and variance quite a bit. However, there are some in academia that believe there is no such thing as luck (Hales & Johnson, 2014, p. 520). They believe that variance always equals out over time and, thus, luck does not exist. However, what they are really talking about when they say luck does not exist is luck over the long term with a large sample size (Hales & Johnson, 2014, p. 520). They are refuting the idea that some people are just luckier than others and will come out ahead even after normal variance balances out. Because the short term luck is really variance manifesting itself, they declare that luck does not exist (Hales & Johnson, 2014, p. 520). However, there are times in sports where a coach, player, or team may not be able to wait out a bad turn that can be accounted for by variance. I have already discussed the fact that NFL teams can finish more than 2 wins different than their on-field performance might have been expected to yield (Barnwell, 2013). While this can go either way, the timing of the swing could potentially have a great deal of impact on a coach or a franchise.

To illustrate an example of how this could come to pass, take the following two coaches. The first coach has been with his team for many years. He has won a fair amount of games and been to the play-offs more than once. However, in recent seasons the team has grown stale and his job is now in jeopardy. In the upcoming season, it is widely believed that he must reach the play-offs or he will be fired. His team is good and performs to a level where they should win 10 games. Because of the variation observed, measured by the PWE and actual wins differential, that team could finish with 8 wins, miss the play-offs, and the coach could be fired. On the opposite end of the spectrum, the team could win 12 games and be a high seed in the play-offs which would likely buy the coach some time with the team (Barnwell, 2013).

The second coach has been with his team for just two years. He took over a team that was poor, but has improved their win total from 2 in the first year, to 5 in the second year. This year they perform to a level that should yield them 7 wins (Barnwell, 2013). After factoring in variation, this team could end up with another 5 win season, which would be 3 straight well under .500 performances, and show stagnation in development. This could lead to the coach being fired and, at the very least, it would put him under a great deal of pressure. On the other hand, they could win 9 games and be on the cusp of earning a play-off berth (Barnwell, 2013). This would suggest that the coach is successfully building the team and there would be reason for optimism.

In reality, both teams are likely to win the amount of games that their on-field performance deserves and the decision on whether or not to retain them would be less cut and dry (Barnwell, 2013). However, I believe these examples show how different the narrative could be through no fault of the coach or team. While these coaches both have

a large enough sample size that their entire body of work should be relatively close to what their teams have earned, both coaches are at a critical point in their careers where one season has a great deal of impact. Because one short stretch has such an impact, variation can also have a great deal of impact. It certainly appears reasonable that this could happen. This is the luck that I am referring to in this study. It isn't the long term luck that may not even exist, but the short term variation that occurs at just the wrong point to have an outsized impact.

It certainly appears that luck has an impact on some team's results over the course of an NFL season, but this could still be accounted for by problems with the PWE as opposed to luck. While I previously discussed why I don't think that is the case from the perspective of the PWE being a good predictor that is difficult to improve upon, it may also be useful to examine how much luck factors into the NFL from the perspective of how the actual distribution of win totals compares to a league that was completely based on luck. Brian Burke of Advanced Football Analytics looked into exactly that (2007). He found that a league based purely on luck, where every team had exactly a 50% chance to win each game, would result in a perfect bell curve of team win totals over time. 8 wins would be the most common win total and almost no teams would win all of their games or none of their games. This was used as the pure luck league because the results mimicked the results if you flipped a coin to decide the outcome of each game. However, this would also be the result if the league had perfect parity and each team was exactly as good as each other team. I know of no reason to believe that is the case so we can proceed using it as a representation of a pure luck league.

Burke (2007) ended up plotting the actual team win totals along with the expected

win totals of a purely skill league and putting them all in a graph. From the graph it was evident that the actual outcomes were a blend of skill and luck. This conclusion was supported by Burke's chi-square analysis and his further plotting of combinations of luck and skill in varying combinations such as 10% luck, 90% skill, 20% luck, 80% skill, etc. At 52.5% luck, his curve became statistically indistinguishable from the actual team win total curve. This suggests that the outcome of an individual NFL game is decided by luck 52.5% of the time. Considering that the better team would win half of those games, pure luck being 50/50, it suggests that the better team wins only 74% of the time in the NFL.

Burke's work is a perfect example of the excellent analysis being done on commercial websites as discussed above. The techniques used to investigate an important question would be right at home in any academic paper investigating a similar question. It is difficult to find much lacking in his approach or his thoroughness. Given that Burke has been formally consulting with NFL teams since 2011 ("About Brian Burke", 2015, p. 1) teams at the highest level seem to respect his research as well. This all would suggest that his work has merit. This suggests that luck has a fairly strong influence on the outcome of games in the NFL. For our purposes, however, the amount of effect that luck has is less important than that it does appear to be a significant factor in the outcome of games.

This suggestion that the outcomes of NFL games are significantly influenced by luck may explain an issue discussed above with the PWE. Previously it was noted that the PWE struggles to accurately predict win totals for teams that win almost all of their games or almost none of them. This could have suggested a weakness in the formula itself. However, in light of Burke's (2007) work, it is entirely possible that any NFL team

that achieves such a record does indeed experience the good or bad luck that the PWE suggests. If the better team only wins 76% of the time, it would take a great team with great luck as well to go undefeated (Burke, 2007). The history of great NFL teams that lost a game or two and the history of terrible teams that managed to win games would certainly support that idea.

This suggests luck can be a factor both from the PWE side and the win total side and provides a basis for the study. From this basis, it is reasonable to ask the question, if luck is a factor in results, how much might it factor in coaches getting fired? I have previously discussed how team performance is not the only factor in deciding whether a coach is fired, but it is one of the main factors. If luck has a hand in one of the main decision making factors, it stands to reason that it would be a factor in the decision itself. Whether any of the management teams in the NFL are aware of that and accounting for it is unknown.

As noted above, teams who finish with a high differential between actual wins and the number of wins the PWE would have predicted are likely to regress towards the PWE the next season (Barnwell, 2013). There are also other measures that function in a similar manner. One of those measures may be the ability to win close games (Barnwell, 2014). It has been common wisdom in sports for a long time that some teams simply perform better under pressure than others (Barnwell, 2014). However, the idea that some teams are better in close games does not seem to be true when examined with statistical rigor (Barnwell, 2014). Bill Barnwell, the same writer who provided some of the PWE analysis referenced above, examined the records of teams in close games and found a similar regression to the mean for both teams that won a high percentage of their close

games, defined as games decided by 7 points or less, in a given season and those who won a low percentage (2014). While individual teams may avoid this regression for a season or two, due to an extremely small sample size, when you group these teams together, the regression is apparent. As groups, both end up very near 50% the next season. This is a perfect example of variance evening out as the sample size grows. As noted above, a 16 game regular season does not appear to be a large enough sample size for variance to reliably even out. When you consider a subset of that season, in this case close games, variance is even more of a factor. This could explain why some teams have runs of multiple years where they either win a high percentage or lose a high percentage of close games. If a team only plays a small number of these games a year, it will take many years, in some cases, for variance to balance out (Barnwell, 2013). Because professional sports rosters experience near constant change, by the time this luck changes, either for the better or worse, the team may look a great deal different. This may have led to the idea that some teams were better in close games than others, but it appears not to be the case.

If this is not the case, it is another aspect of professional football that suggests luck is a significant factor. When I discussed the PWE differential earlier, I mentioned that the possibility existed that the differential was not due to luck, but due to flaws in the formula (Luo, 2014). Certain teams could perform in such a way that they are outliers when measured by the PWE. The suggestion that most of those outliers regress to the mean the next season casts doubt upon that possibility, but it cannot be completely dismissed (Barnwell, 2013). This further suggestion, that close games have a similar regression pattern and teams are unlikely to be significantly better or worse than others in

close games, casts further doubt on outliers being due to something other than luck represented by short term variation.

This pattern continues when we take a look at another measure that can predict success in the NFL, turnover margin. Kevin Rudy conducted an analysis of turnovers, how they impacted the NFL season, and how many turnovers can be accounted for by luck (2014). He found that, as expected, turnover differential is a significant factor in both a team's winning percentage and its scoring differential. While this is to be expected, it's important to note given the number of things that used to be conventional wisdom that don't hold up to statistical tests (Barnwell, 2013). Rudy (2014) continues his analysis by examining fumbles caused and recovered by a defense. He found that luck played a huge role in determining both. He went on to test interceptions and again found that random chance, or luck, played a huge role in determining the number of interceptions a defense got. In fact, the only aspect of turnover ratio that Rudy found to be positively correlated from the first half of the season to the second was interceptions thrown by the offense.

If a team only has control, and only some control at that, of one aspect of the turnover battle and winning the turnover battle has such a drastic impact on winning, it again suggests that luck has a real impact on the results of games in the NFL (Rudy, 2014). Rudy's (2014) study looks at one season only to mitigate the impact of personnel and coaching changes on the results. It does not account for schedule differences, however, which could, potentially, be a factor. Regardless, getting these types of results across the league make it unlikely to be skewed too greatly. This becomes just another piece of evidence that suggest that luck has a significant impact on the NFL. It looks

more and more likely that the outliers seen by examining the PWE differential are, in fact, teams that have experienced extreme luck one way or another be it through turnover margin, record in close games, or another measure that has yet to be uncovered.

CHAPTER THREE: METHODOLOGY

In order to determine the effect that luck has on coaching retention in the NFL a model can be built using previously discussed relevant factors, such as PWE, to put through a binary logistic regression. Using a model where the dependent variable is equal to 1 if the coach was fired at the end of the season or 0 if he wasn't should tell us what the relationship is between luck and coaching retention. It should also inform on how that relationship compares with other factors. A binary logistic regression can be used to predict a categorical, dichotomous variable, such as whether or not a head coach was fired, using a set of predictive variables (Wuensch, 2015, p. 1). By identifying the major factors that could be used to determine whether or not to fire a head coach and determining a variable to measure each one, it can be determined which factor is the most predictive for whether a coach is fired. For the purposes of this thesis, the standard for significance will be set at .10 (Orzel, 2011).

One binary logistic regression run on the model presented below could be useful in finding whether luck has an effect on coaching retention in the NFL. However, there are good and bad luck effects present at the same time in the data set. Because of this, it may be useful to isolate good luck and bad luck in turn and run a regression of the same model on each side of luck by itself. After the initial regression, the data set will be split into two subsets. The first will include only bad luck teams or, stated another way, teams

who won fewer games than the PWE predicted they should. The second will include only good luck teams which are teams who won more games than the PWE predicted they should. The same model will then be run against each subset and the results presented. This should give some idea as to whether one side of luck, good or bad, is more powerful than the other. Combined with the original regression, there should be a clear picture of what, if any, effect luck has on coaching retention in the NFL and how that could be used to make better decisions when deciding to retain a coach or not.

3.1 Building the model

Having established that the PWE appears to be a good measure of the number of wins a team's performance should merit, given average luck, and that the difference between the PWE and a team's actual winning percentage should represent the luck experienced by a team in a given year, it should now be possible to build a formula around that idea and test its merits. To test how much luck factors in to whether a coach gets fired or not, we must test as many factors that could influence the decision as possible. Including as many as possible will allow comparisons between each factor and make it clear whether luck has a significant effect.

In order to test multiple factors and how much they effect whether or not a coach is fired, I will use binary logistic regression. Whether the coach was fired or not will be the binary dependent variable and each of the factors that may have an influence on that decision will be independent variables. This is a good fit for the question at hand. Because a logistic regression allows for the examination of the influence of various factors in a dichotomous outcome, it should provide a strong indication of whether luck is

having an effect (Anderson, n. d., p. 1). In order to do that, however, it is imperative that the proper factors are included in the analysis and that their representative variables measure those factors in the right way to answer this question. In order to build the formula in that way, each factor that is to be included must first be considered.

The first factor that I want to include is one of the most basic. At the highest levels of sport, winning games is the ultimate measure of performance. If a coach is measured by his team's results on the field, I would expect the team's winning percentage to have a strong effect on whether the coach is fired or not. Because the PWE formula results in an expected winning percentage, my formula will use a team's winning percentage to represent their actual results on the field for a given year.

The next factor to consider is the PWE. As previously discussed, the PWE is a measure of predicted wins based on a team's performance on the field had their luck been neutral. In this way, the PWE is a measure of a team's on field performance, as opposed to the traditional win percentage which is a measure of results. The difference between performance and results is, ultimately, the basis for this study.

As noted above, what is most interesting here is the luck factor which is, at its core, measured by the difference between the actual winning percentage and the PWE predicted winning percentage. Therefore, the model constructed here contains the difference between the PWE and the actual win percentage of a team for a given season, labeled as "Diff16" in all tables. This puts a measure of luck directly into the model. In order to make the results of the logistic regression simpler to interpret, the variable representing the difference between the two will be multiplied by sixteen. This will give us the luck variable in the number of wins it represents rather than a decimal number

representing the winning percentage. This will make the findings of how much a one unit change effects whether a coach is fired more easily digestible.

While winning percentage and the PWE give us measures of success for a team, neither feature a set standard for what is or is not a successful season. It is left to each franchise to determine how many wins is enough in a given year. However, the NFL does feature a line of demarcation each year between successful teams and unsuccessful teams. The champions of each division, along with the two best records among the rest of the teams in each conference, make the play-offs. This can act as a sort of default measure of success for a franchise. This is not a perfect measure as the top twelve teams by record don't always make the play-offs, but making that field and having a chance to win a title may have an impact on whether or not a coach keeps his job. Therefore, the model needs to include it. Because this measure is being used as a demarcation line between success and failure, it makes sense to use a binary variable, labeled as "Play_Offs" in all tables. While some attempt could be made to account for additional success or failure in the play-offs, that would be beyond the scope of this particular variable and model.

The next potential factor in the decision to change coaches that I will include in this model is the tenure of the current head coach. In the discussion above, it was found that some studies suggested changing coaches made things worse while in other studies it was found that changing coaches was the right decision for teams that were not performing. While the conflicting findings will not be settled here, it does suggest that tenure is at least worth investigating as having an effect on the coaching change decision. Adding to that the potential for a long tenured coach to have more political power than a

shorter tenured one and the possibility exists that tenure is a significant factor. Measuring tenure for this study is relatively simple. For this model, the variable measures the number of years the current head coach has been at his current post, labeled as “Tenure” in all tables. For first year head coaches, the variable value will be a one. Each year the coach is retained, the value will increase by one.

The number of years a head coach has held the position is a simple way to address tenure, but there are a number of other related factors beyond that which may be informative as to why a coach was fired or retained. For instance, the tenure variable used here does not address the remaining years of a coach’s contract or the financial implications of firing him at a certain time. Firing a coach whose contract is up at the end of the season is much more palatable than one who has many years left and would be paid a great deal of money if not retained. Beyond such quantitative factors, political factors, as previously discussed, can have a significant impact on the decision. If a coach has a strained relationship with management, he is more likely to be fired. All of these things work in concert with other factors discussed here to inform the final outcome. However, it is difficult to gather accurate information for some of these factors, such as coaching contract details and others are far too subjective to include, such as the relationship between the coach and management. Because of that, the simple tenure variable will be used here. Further studies could be conducted to investigate additional factors.

The last aspect of the decision making process that must be addressed in this model is the expectations of the franchise. As noted above, success for an NFL team is not a uniform measure. Some teams start the year simply hoping to be competitive, gain

experience, and build towards the future. Others start the year feeling that anything short of a Super Bowl victory will be a failure. This can make judging what gets a coach fired quite difficult. In order to mitigate the effect the varying expectations might have on the model, the model must make an effort to measure what those expectations are likely to be. While it is impossible to know for sure what management expects from their team in a given year, there are two things that may give us a good idea.

The first is how the team has performed in prior seasons. This is important for a couple of reasons. The short term history of a team may give a good indication of where the team is from a talent and development stand point. Teams that have won a lot in the last two to three seasons might expect more than those who have been suffering through losing campaigns. Taking a longer view, franchises who have won a lot in the last five to ten years, or even longer, may tend to have higher expectations than those who are, traditionally, near the bottom of the league. Holmes (2011) investigated this theory on college football coaches and found it to be true. His findings are informative as to how to structure the measure we use to represent expectations in the model. He used win percentages from one year ago, two years ago, three to ten years ago, and eleven to thirty years ago. The three to ten years ago average was the strongest predictor of the group. The variables for last season and two seasons ago had some predictive strength, especially last season, but individually they just aren't strong enough.

For my model, I'd like to avoid having several different measures of past win percentage. For one, I'd like to group those that have an effect together so they are as strong and accurate as possible. For another, adding multiple win percentage variables from a similar time period may cause issues of collinearity. While they would not be

directly related, winning teams in the NFL may tend to keep winning, while losing teams may tend to keep losing. I'd prefer to avoid those issues all together and identify one time period to use. To that end, Holmes would suggest factoring in the last ten seasons. However, there are some fundamental differences between the NFL and college football that make that a questionable decision. As previously mentioned, the salary cap and rule structure of the NFL encourages parity. This is in stark contrast to college football where some schools pour much greater resources into their programs than others. This could lead to much quicker changes in fortune in the NFL. Because of this, the history that is applicable to expectation levels may be less. For this reason, I will be using the past five seasons win percentage to measure expectations, labeled as "Last5" in all models. As with the differential variable, converting the win percentage to the number of wins per season it represents will give us more easily digestible output so the percentage will be multiplied by sixteen.

The other factor that may give us insight into what the expectations may have been for a given team is what sort of trend their winning percentage had seen in the past five seasons. This may help to address some of the problems that averaging out winning percentages over several years could have. For example, if the model only used the past five seasons winning percentage, the following two teams would look exactly the same. Team 1 won 10, 8, 3, 4, and 6 games over the last 5 years. Their coach got fired after the 3 win season and the new coach has increased their win total each of his first two seasons. Team 2 won 3, 4, 6, 8, and 10 games. Their coach has been in place for all five of those seasons and finally got them into the play-offs last season. Team 1's expectations are likely lower than team 2's. While team 1 is likely looking for another

increase in their win total and maybe a play-off berth, team 2 is likely looking to win their division and contend for a Super Bowl spot. Without identifying the trend, it would be impossible to tell the difference between the two.

In order to measure the trend, the model will simply use a value that reflects how many seasons in a row a team’s winning percentage has moved in the same direction, labeled as “Trend” in all models. Positive values will be used for teams with a rising win percentage and negative values will be used for those who have falling win percentages. Because we are using five seasons of history for the winning percentage piece of expectations, the trend will only reflect the past five seasons as well. The same reasons given above as to why history beyond five years is unlikely to have much impact apply. This variable will only reflect the current active streak for a given team. If they have improved their win totals for the last 3 seasons they will have a 3 for this variable. If they have lost more games during the last 2 seasons, they will receive a -2.

Table 1. Model Variables

Name	Definition	Relationship
Diff 16	The difference between the PWE expected winning % and the actual winning percentage multiplied by 16.	Negative
Last5	The average of the last 5 seasons winning percentage.	?
Tenure	The number of years the head coach has been in his current position.	?
Play_Offs	Whether or not a team made the play-offs for that season.	Negative
Trend	How many years in a row a team has won (or lost) more games than the previous season. Only the current active streak is included.	?

Table 1 identifies the variables that will be used to measure the factors identified above and declares what, if any, relationship is expected between the variables and whether a coach is fired. I expect that Diff16, the variable used to measure luck, and Play_Offs, whether or not a team made the play-offs, will have a negative relationship while the others could go either way. This is simply due to good luck and making the play-offs both indicating that a team had a better season than those that had bad luck or didn't make the play-offs. Better performance on the field should lead to less firings.

Less clear is the relationship with Last5, the average of the last 5 season's winning percentages, Tenure, the number of years a coach has been in his current position, and Trend, which is the number of years in a row a teams win total has moved in the same direction. Each of these variables has forces pushing them in different directions. Because of that, they may be positive or negative relationships with the dependent variable.

The model defined above is as follows:

$$(Coaching\ Retention) = a + b_1(Diff16) + b_2(Last5) + b_3(Tenure) + b_4(Play_Offs) + b_5(Trend)$$

In this model, coaching retention, or whether a coach is fired or not after a given season, will be predicted by Diff16, Last5, Tenure, Play_Offs and Trend. The coefficients for each variable should give some clue as to the relative predictive power of each factor in the given formula.

3.2 The data set

The model that has been identified to this point should do a good job of telling us if luck is a significant factor in whether or not a coach gets fired in the NFL. In order to answer that question, a good data set must be identified to run the model against. With all of the variables previously identified, gathering the data was relatively easy. Deciding on a sample size and parsing through a couple of issues inherent in the data set was slightly more difficult. As for sample size, as previously discussed, the NFL is a rapidly evolving league so historical trends that date back too far may not be especially informative. On the other hand, this sort of analysis lends itself to a longer time frame than information that may be used in deciding to fire a coach. Weighing those two factors, using the last 20 seasons, starting with 1994 and ending with 2013 seems to be a good balance. At the time of compilation, the data set was only complete through 2013 which is why 2014 was omitted.

The data set was compiled using Pro-Football-Reference.com. In addition to the basic information required to compute the variables included in the model, there were some adjustments to be made that required interpretation of certain situations. The first such situation was how to handle coaches who retired or otherwise left teams without being fired. Because the aim here is to examine the decision on whether or not to fire a coach, the only coaching changes that are relevant are those where a management team decided to go in another direction. For that reason, the data set reflects a 1 for the coaching change variable only whether a coach was fired or not with all other outcomes, such as returning the next season or retiring, being a 0. This requires some judgments to be made. It is sometimes difficult to determine when a coach is fired and when he has

chosen to walk away. Especially in cases where the coach leaves to take another coaching job, it can be difficult to determine if he left to avoid being fired. Every effort was made to determine what most likely happened in any given situation, but the determinations made here are not infallible.

The second situation was what to do with teams that didn't have five years of history to fill the average winning percentage aspect of the expectations factor. This applied to expansion teams such as Jacksonville and Carolina, but also to the new version of the Cleveland Browns. For teams that moved, the franchise simply continued in a new location so the data continued on as if nothing had happened. For Cleveland, their franchise moved to Baltimore and continued there so the expansion team they were awarded was a new start. While there are a number of ways to deal with missing data, as there would have been for these teams until they built enough history, it was ultimately decided to remove them from the data set. The potential that expansion teams may operate a bit differently from other, more established, franchises cannot be discounted. Trying to measure them as you would an established franchise may introduce problems. For that reason, it seemed safer to remove their first five years from the data set. The sample size is still plenty large enough without those seasons. With those issues resolved, the remaining data set has 610 cases. Running a binary logistic regression with the identified model on those 610 cases will identify if luck is a factor in whether or not a head coach is fired in the NFL.

The data set includes all cases, but it is useful to run the same model with only good luck cases and only bad luck cases each in turn in order to isolate the two. After running the model on the entire data set and examining the results, the same model will

be run on only the good luck cases and then only the bad luck cases in order to see if one direction of luck is stronger than the other. This may give us insight into whether bad or good luck is more predictive of coaching retention which could shed some light on just what management teams are doing now and might need to look out for in the future.

CHAPTER FOUR: RESULTS

With the model and methodology defined, it is useful to look at some descriptive statistics for the data set before proceeding to the three regressions described above. Of particular interest here is what causes a head coach in the NFL to get fired. If that can be understood, it may be of use when looking at the results of each regression and answering the question of what, if any, effect luck may have on head coach retention in the NFL.

4.1 Data set analysis

Table 2. Dependent Variable Frequency

Value	Frequency	Percent
0	477	78.20%
1	133	21.80%

Of the 610 cases in the data set, the coach was fired 133 times or 21.8% of the time. This amounts to a little more than 6 coaches fired each NFL season. 78.2% of the data set is made of up of cases where the coach was not fired.

Table 3. Descriptive Statistics – Variables

	Mean 0	Std Dev 0	Mean 1	Std Dev 1
Tenure	3.828	3.113	4.15	2.781
Wins	8.665	2.795	5.684	2.645
Diff16	0.183	1.205	-0.483	1.224
Avg Wins – Last 5	8.103	1.935	7.621	1.872

The average tenure of a coach that was fired, represented by Mean 1 in table 3, was slightly higher than the average tenure of coaches who were not fired, Mean 0 in table 3. Only 13 of 135 first year head coaches were fired, which is a little less than 10%. That rises to 27 of 120 in the second year which, at 22.5%, is higher than the overall average. The number of wins is substantially higher for coaches that were not fired compared to coaches who were. Also, the average luck of coaches who were fired was much worse than those who were fired. Coaches who were fired averaged almost half a win less than the PWE predicted. Finally, the average number of wins over the last five seasons is higher for coaches who were not fired, but by less than half a win per season

Table 4. Descriptive Statistics - Cases

	# of Cases	Coach Fired	%
Made Play-Offs	235	14	5.96%
Win Total Trending Up	276	62	22.5%
Last 5 Yrs Avging 10+ Wins	104	17	16.3%

Making the play-offs substantially decreases the odds that a coach is fired after a given season, dropping the chances to less than 6%. Having a win total trending up compared to previous seasons does not appear to have much, if any, effect. The overall look at any positive trend actually has a higher incidence of coaches getting fired than the entire data set does. A closer look shows that only a two-year trend in a positive direction, meaning that a team won more games last year than they did the year before,

then won more games this year than they did last year, has a lower incidence of firings than the overall average. Even then, approximately 17% of coaches are fired. The win total trend doesn't seem to be a good measure.

Consistently winning games, defined in this case as averaging 10 wins or more over the past five seasons, only drops the chances of a coach getting fired to 16.3%. This seems in line with the above finding that the average number of wins over the past 5 years is not much higher for coaches who are not fired as compared to those who are. It is a factor, but far less important than current year performance.

One of the most significant findings here is that the average bad luck team loses more than half a game, on average, more than the average good luck team. This strongly suggests that luck is a significant factor in whether or not a coach is fired. However, more than one piece of evidence is necessary to build a strong case. Investigating the findings of the regressions could now provide additional evidence or refute this finding.

4.2 Regression of the model

The first regression run was the constructed model against the entire data set, including all cases, both good and bad luck. The goal here was to address one of the overall goals of this thesis which is to determine how much, if any, effect the impact of luck has on the retention of NFL head coaches. This regression should help to further answer the question of how much, if any, effect the impact of luck has here.

Table 5. Regression Results – Model

	B	S.E.	Sig.
Diff16	-.30	.09	.001
Last5	-.15	.06	.016
Tenure	.13	.04	.001
Play_Offs	-1.76	.32	.000
Trend	.00	.08	.981
Constant	-.19	.44	.675

Table 6. Dependent Variable Frequency

Value	Frequency	Percent
0	477	78.20%
1	133	21.80%

Table 7. Overall Descriptive Statistics

Variable	N	Mean	Std Dev	Min	Max
Tenure	610	3.90	3.04	1.00	16.00
Last5	610	.50	.12	.19	.83
Trend	610	-.04	1.36	-4.00	4.00
Diff16	610	.04	1.24	-3.32	3.76

After performing a binary logistic regression, where the dependent variable was a 1 if the coach was fired after the season and a 0 if not, the model constructed above was found to have a Nagelkerke R squared of .20. This means that 20% of the variation is explained by the model. While this is not overly high, it is far enough above zero to indicate that the model may have predictive value (“FAQ: What are pseudo R-squareds?”). Given the number of factors that go into the decision to fire a coach in the NFL, and the number of those factors that cannot be quantified, it should not be overly

surprising that the r square measure is a little low. This should not mean the entire model is discounted.

This regression finds luck, as represented by the Diff16 variable to be a significant factor and the second strongest predictive variable to Play_Offs, which was also found to be significant. Both of those variables had a negative relationship with the dependent variable. Coaches that make the play-offs being less likely to be fired is precisely what was expected. That being confirmed here in a model where it is suggested that better luck also makes a coach less likely to be fired further builds the case that luck is a real factor here. It was expected that Diff16 would have a negative relationship and it does.

The Last5 variable was also found to be significant and have a negative relationship with the dependent variable, though it was less predictive than Diff16 and Play_Offs. This suggests that the rise in expectations for a team that has been winning outweighs the good will engendered towards the head coach by past victories. The variable for tenure was found to be significant and had a positive relationship with the dependent variable. This suggests that coaches are more likely to get fired the longer they spend in a given position. Finally, the trend variable was not close to significant. It appears that it is either not a factor or a better measure needs to be constructed.

This initial regression again points to luck being a real factor in whether a coach is fired or not. It is important to note that the minimum, -3.32, and maximum, 3.76, values found for the Diff16 variable represent a difference of over 7 wins. This is a substantial finding suggesting that luck has a powerful effect on the final outcome for a coach. While the minimum and maximum values are relatively close in their absolute

values, it is possible that good luck or bad luck is more predictive than the other. In order to test for that, the two opposing luck forces must be isolated and regressions run against the isolated data.

4.3 Regression to isolate bad luck

In order to isolate bad luck from good, the next regression will be run using the same model, but against a data set that contains only those original cases where the Diff16 variable was negative. This should provide insight into the power of bad luck and, when compared to good luck and the overall regression, may be informative as to whether it is more, less or equally as powerful.

Table 8. Regression Results – isolated for bad luck

	B	S.E.	Sig.
Diff16	-.31	.18	.087
Last5	-.15	.08	.060
Tenure	.14	.05	.006
Play_Offs	-2.05	.55	.000
Trend	-.06	.10	.566
Constant	-.24	.60	.684

Table 9. Dependent Variable Frequency

Value	Frequency	Percent
0	205	69.97%
1	88	30.03%

Table 10. Bad Luck Regression Descriptive Statistics

Variable	N	Mean	Std Dev	Min	Max
Tenure	293	3.70	3.03	1.00	16.00
Last5	293	.50	.12	.21	.80
Trend	293	-.07	1.34	-3.00	4.00
Diff16	293	-1.01	.73	-3.32	.00

The results of the regression to isolate bad luck look remarkably similar to the original model with one major difference. The Nagelkerke R squared is .16 which is slightly less than the full regression. The coefficients and odd-ratios for the variables are similar to what was found in the previous regression. Diff16, Play_Offs and Last5 remained significant with negative relationships while Tenure was significant and had a negative relationship. Trend was still found to be insignificant.

That the coefficients, significance and relationships did not change much, if at all, suggests that bad luck has a similar effect as luck overall. The magnitude of that effect also appears to be similar and it does not appear that any of the factors in the model are substantially more or less sensitive to bad luck than luck overall. This makes it likely that good and bad luck have similar powers, but the model will be run again, isolating good luck this time, to be sure.

4.4 Regression to isolate good luck

In order to compare the power of good luck to that of bad luck, the same regression, using the original model and, this time, using a data set with only good luck cases, was run. Comparing the results of the two should be informative as to the relative power of the two opposing directions of luck.

Table 11. Regression results – isolated for good luck

	B	S.E.	Sig.
Diff16	-.29	.27	.286
Last5	-.17	.11	.116
Tenure	.12	.06	.057
Play_Offs	-1.60	.40	.000
Trend	.09	.12	.477
Constant	-.10	.74	.893

Table 12. Dependent Variable Frequency

Value	Frequency	Percent
0	272	85.80%
1	45	14.20%

Table 13. Good Luck Regression Descriptive Statistics

Variable	N	Mean	Std Dev	Min	Max
Tenure	317	4.09	3.05	1.00	15.00
Last5	317	.50	.12	.19	.83
Trend	317	-.01	1.38	-4.00	3.00
Diff16	317	1.00	.72	.00	3.76

As with the regression isolating bad luck, the results here are similar to what was previously found. The Nagelkerke R squared is .16 which is identical to the bad luck regression. The odds ratio for the differential variable here is slightly higher than the first two regressions. While this does line up with an ever so slight difference in effect, bad luck being slightly stronger than the overall, good luck being slightly weaker, it isn't nearly definitive enough to draw conclusions from. That Diff16, Play_Offs and Last5 continue to be significant with a negative relationship and similar coefficients further

suggests that the two sides of luck are very close to the same in predictive power. Tenure and Trend were also found with the same significance, relationship and similar coefficients. All three regressions generate similar enough output that it suggests good and bad luck have similarly powerful effects.

CHAPTER FIVE: DISCUSSION

This research set out to meet two objectives. The first, to gain an understanding of the impact of luck on an NFL team's outcomes during a given season. The second, to determine how much, if any, effect the impact of luck has on NFL head coach's retention. After reviewing other research, it became clear that luck has a real impact on the outcomes of individual games in an NFL season (Burke, 2013). There are some statistical measures, such as record in close games and turnover margin (Barnwell, 2013), that can help identify when a team may have been lucky or unlucky over a period of time, but they lack the ability to specifically quantify the end result of luck on a team's record.

To get to that specific measure, the Pythagorean Win Expectation was used to establish a baseline (Luo, 2014) of what a team's record should have been with neutral luck. The difference between a team's actual record and the PWE predicted record was then found and used as a measure of the effect of luck in a given season. This quantified the impact of luck on an NFL team's outcomes during a given season.

With that objective achieved, it was on to determining how much, if any, effect the impact of luck has on NFL head coach's retention. A model was built with variables accounting for luck, tenure, recent performance, whether a team made the play-offs and whether a team was trending towards winning more games or less over previous years. This model was then put through a binary logistic regression in order to investigate the impact of luck on NFL head coach's retention. The results indicate that luck is a significant factor in deciding whether or not a head coach in the NFL is fired or retained.

5.1 Data set analysis discussion

Before further discussing how luck is a factor in whether a coach gets fired, it is important to understand why coaches get fired. Taking a deeper look at the data set used for this study can help with that aspect. The understanding that just over 20% of cases in the data set represent seasons that ended with a head coach getting fired gives some idea of the magnitude of the situation and the likelihood that this will happen. While most head coaches will retain their job each year, having approximately one in five turn over each season suggests that most won't survive for too long.

The difference between coaches who get fired and those that do not are sometimes subtle and sometimes obvious. The average tenure of the two groups is fairly close, but those that get fired tend to have slightly more years at their current job. That is largely explained by coaches in their first year being less than half as likely to be fired when compared to all other coaches. There seems to be little protection after the first year. One thing that does seem to protect coaches, no matter what their tenure is winning games this season.

Coaches who did not get fired averaged almost three more wins over the current year than their fired counterparts. This, combined with the finding that coaches who made the play-offs were fired less than 6% of the time suggests that current year performance is the most important thing. This also supports the idea that luck is a substantial factor as luck directly effects the number of wins a team achieves in a given season and, in certain situations, could decide whether or not a team makes the play-offs. Coaches who are fired average almost half a win less than their PWE while coaches who are retained average almost two-tenths of a win more than theirs. This is an important

finding because it directly suggests that coaches who are retained are luckier than those who are fired. The luck factor does not account for the entire difference between the number of wins each group achieved, but it does account for some of it.

Current year success appears to be important, but success in the recent past seems less so. The difference between the average number of wins over the last five seasons for the group of coaches that was retained was a little less than half a win per season more than those who were fired. While there is a difference, it is much closer than the difference in current year wins. This may, again, be the conflicting forces of wins in previous seasons building up good will towards the coach, but also building up expectations that must be met. Regardless of why it is, the effect is clear. Winning in previous years is good, but if a coach doesn't get the job done in the current season, he may not get another chance.

Despite the suggestion that winning this year is the most important thing, there are some suggestions that it isn't the only thing having an effect. The fact that some coaches get fired even after making the play-offs and get fired even after averaging 10 or more wins over the past 5 seasons suggests that even winning won't always save a coach's job. As previously discussed, inflated expectations or power struggles in the front office could factor in here (Breer, 2013), but it is difficult to know exactly why each decision is made. The most important finding to note from the data set is the substantial difference between the luck of coaches who are retained and those who are not. This further suggests that luck plays a substantial role in the final decision.

5.2 Initial regression discussion

Analyzing the data set again suggests that luck is a substantial factor here. If that is true it should be evident in the regression results as well and it is. A 26% reduction in the likelihood of getting fired for a one unit increase in the Diff16 variable, in this case that would be one additional win over a season, is significant. Of the 610 records in the data set, 60 of those finished a season more than 2 wins higher or lower than the PWE would have predicted. This is almost 10% of the data set which is equivalent to about 3 teams per season. For these teams, luck could have a serious impact on whether or not a coach is retained. The effect may be more subtle for the rest of the league, but it may still sway the final decision under the right circumstances.

The variable for the average number of wins in the past five seasons indicates that more wins over recent seasons makes a coach less likely to be fired. This is at odds with the suggestion that more wins result in higher expectations which, in turn, results in a coach being more likely to be fired. Instead, it would appear that the power of a better recent track record outweighs any raised expectations from management. This could indicate that coaches can earn some goodwill with prior good performance that can overcome a bad season. As discussed above, there are a number of other factors involved with tenure that could also be influencing this finding. More research is necessary to evaluate which factors have the most influence. The tenure factor is less powerful than luck, however. This could be due to the conflicting powers at work with this variable, but it also could indicate that luck is a significant factor in the overall decision making process.

The tenure variable finds that a coach is more likely, on average, to be fired as his tenure advances. This seems to be at odds with conventional wisdom that a coach becomes more secure the longer he stays with a team, but this could be, again, the result of conflicting powers at work in this variable. The powers, in this case, would be the initial grace period that many new coaches get before they are likely to be fired against the job security that a long tenured coach may have. . If most coaches get a three year grace period, they would be much more likely to be fired after their third, fourth, or fifth season than their first or their second. If there are a lot more of these coaches than those who stay at their jobs long term, which seems plausible on the surface, you could have a case where gaining tenure would be detrimental to your chances of keeping your job. More research could be done here in future studies, but structuring the tenure variable to distinguish between those early in their tenures as head coach, and thus potentially enjoying a grace period, and those who were well past any grace periods may be informative.

The variable indicating whether a team made the play-offs was the strongest predictor in the model. An 83% reduction in the chance a coach will get fired is significant, if not overly surprising. This holds with conventional wisdom that making the play-offs is seen as success for most NFL head coaches. Considering that good luck will raise a team's win total and raise its chances of getting into the play-offs, the effect of luck may have some impact on the play-offs being such a strong predictor.

Overall, analyzing this model seems to suggest that luck has a real impact on whether or not a coach gets fired. However, it does not distinguish between good and bad luck. It could be the case that the luck effect on the head coaching question is more

pronounced one way or the other. Put another way, bad luck could be more likely to get you fired than good luck is to save your job or vice versa. This is easily testable with a slight adjustment to the data set. To isolate bad luck and good luck respectively, two further regressions will be run. The first will be run using only cases with a differential of less than zero. This will isolate the effect of bad luck. The other regression will be run using only cases with differentials greater than zero. This will isolate the effect of good luck. Comparing the results should reveal if one is much more predictive than the other.

5.3 Discussion of luck regressions

Good and bad luck being equally strong predictors of a coach getting fired makes sense. If the effect seen here is luck, it should be just as likely that a coach on the brink of being fired gets lucky and wins a couple extra games to save his job as it would be for him to get unlucky and lose a couple extra games to lose his job. The overall statistics of the data set bear that out. From the 610 cases in the data set, 317 had good luck while 293 had bad luck. The good luck teams won 58% of their games while the bad luck teams won 42% of theirs. However, when you adjust for luck and look strictly at performance, as measured by the PWE, the two groups look much closer. The good luck teams' performance should have earned them the win in 51.7% of their games with neutral luck. The bad luck teams should have earned wins in 47.8% of theirs. This difference of less than 4% is less than one additional victory per 16 game season.

Table 14. Good and bad luck teams

	# of Cases	Win %	PWE
Good Luck	317	58%	51.7%
Bad Luck	293	42%	47.8%

This finding suggests that winning teams and losing teams both have about the same likelihood of suffering bad luck or experiencing good luck. As discussed above, the PWE is not a perfect formula and, even if it were, the difference between the PWE and a team's actual winning percentage may not be completely explained by luck. If teams that had a high number of wins were more likely to be judged as lucky, it would cast doubt as to whether the effect were truly luck as opposed to a manifestation of something else. In this case, with good and bad luck having similar PWEs, it seems that the majority of the effect could be due to luck.

This difference, in addition to the previous findings that the differential was a significant factor and that good and bad luck are equally predictive, suggests, again, that luck has a real effect on the decision making process. At this point, a good case has been made that luck has something to do with whether a coach gets fired.

5.4 Further Research

There are a number of aspects of the differential variable used here that could be investigated further. The baseline itself, the PWE, may not be as accurate as it could be. Because that formula counts points scored for or against in any game situation as being of the same value, teams are rewarded for running up the score in wins or making potential blow-out losses close after the other team has put in their reserves. Neither of these things is necessarily indicative of how good a team is, but the current formula skews towards teams that take these routes. Further investigation into a formula that weights points scored based on the game situation could be useful to finding a more accurate baseline from which to start. In addition, there could be additional metrics,

beyond points scored for and against, that could be useful in building a better baseline. This could also use further investigation.

The potential for certain teams to consistently show as lucky or unlucky is something else that should be acknowledged and could be the basis for further research. A better baseline may address this to some extent, but there could be some teams which simply perform in such a way that they will always seem to be experiencing more good or bad luck than they should. These teams should be identified and then studied to find out why they don't fit into the current formula well. This could be useful in improving the formula and in the decision making process. Understanding when the differential is a factor of luck and when it may not be is crucial to knowing when to use it to inform your decision.

Also, more research needs to be done to find out how much the performance of a head coach and his team is a factor in the coach keeping his job. As previously discussed, there are many factors outside of wins and losses that influence a management team's decision. Finding out just how influential these factors might be could be informative into deciding how much a change in wins and losses matters overall. If the model with wins used here only explains 28% of variance, then 72% is still unaccounted for. While that is unlikely to be all factors outside of on field performance, those factors could still be a significant influence.

Finally, investigation into whether a sports management team that knew and understood the influence of luck would significantly change their decision making process needs to be done. While it seems that an 8 win team that should have won 10 games but was unlucky would be judged differently than an 8 win team that should have

won 6 games but was lucky, that may not always be the case. While that knowledge informs the management team about on field performance, it may not change the perceptions of fans and sponsors. Whether those perceptions could be shifted by publicizing this sort of data is difficult to know. The debate over how useful advanced metrics are continues in the sports world at this time. There are factions of fans on either side of the debate and the sponsors would no doubt be greatly affected by the reaction of those fans.

5.5 Thinking outside the norm

As previously mentioned, any thinking outside the norm is, at once, a potential breakthrough and a potential disaster. While the potential to gain an edge over the competition is tempting, the potential to alienate fans and lose sponsors is chilling. As time goes on, thinking that was, in the past, outside the norm may become more accepted. While this makes it much easier to implement such thinking, it also eliminates much of the edge to be gained over the competition. In the end, the amount of risk an organization is willing to take in this area is directly proportional to the potential gain they could receive if it pans out. It may take a franchise with little to lose to gamble on a risky, cutting edge philosophy and succeed in order to introduce it into the main stream.

5.6 Conclusion

The research objective to gain an understanding of the impact of luck on an NFL team's outcomes during a given season has been well addressed by other sources. The work of Brian Burke (2013) and Bill Barnwell (2013) in particular made clear that chance has a real impact on winning games in the NFL. What was less clear is just how much

impact that could have on a given team's record. Using the PWE as a baseline, the luck effect for each team and season was quantified and evaluated. The difference between the luckiest team, +3.7 wins, and the unluckiest, -3.3 wins, was substantial. Any factor that could account for a 7 win difference over a 16 game season should be taken seriously.

With that objective achieved, the investigation to determine how much, if any, effect the impact of luck has on NFL head coach's retention could begin. The model to be used for binary logistic regression was built and regressions were run against the entire data set, only those cases indicating good luck and only those cases indicating bad luck. The results again indicated that luck was a significant factor and that both sides of it, good and bad, were similar in their predictive power. That a one unit increase in the luck variable was found to reduce the chances of a coach being fired by 26% is notable.

Luck alone is, ultimately, a significant and substantial factor in determining whether or not a coach gets fired. Analysis of the data set clearly points to luck playing a substantial role. The regressions run here indicate that luck is second only to whether or not a team made the play-offs in predictive power. Because luck impacts the number of games won, it can also impact whether a team makes the play-offs. However, there is quite a bit of variation that is not explained. This is most likely due to the previously discussed idea that the decision to make a coaching change is an extremely complex and nuanced one. There are many factors involved that are not quantifiable and those that are may be judged differently depending on the organization.

It is possible that the league as a whole is failing to account for luck when evaluating a coach's performance or, at least, not accounting for it as much as they

should. If winning games in the current year is the best way to keep from getting fired, that suggests that management teams are focusing on the number of wins and may not be considering how that number was reached. If this is indeed the case, there could be an edge to be gained by NFL organizations in ensuring luck is accounted for during their decision making process. It may not swing the decision in many cases, but it could be informative in certain situations. The fact that fired coaches averaged almost half a win less than the PWE predicted suggests that bad luck is bad for job retention. That could be interpreted as management teams not taking it into account. However, another possibility is that management teams well know the impact that luck can have, but cannot act on it due to outside influences. Sponsors and fans may react to the official standings, necessitating action on the part of the management team. If a head coach loses the support of the fans and sponsors, who play a large role in the financial bottom line of the franchise, he may need to be fired, even if it was caused by bad luck.

Regardless of why luck is a factor in coaches getting fired, it seems clear that it is. If it is clear that luck plays a part in a team getting to their final win total and that regression to the mean is likely for teams that are most effected (Barnwell, 2013), then it is important that sport management teams and everyone involved with sport acknowledge that and understand the implications. Focusing solely on what the results are, regardless of whether your hand is forced by sponsors, fans, or others in power, is failing to take relevant information into account. Failing to take relevant information into account could result in a poor decision being made and the organization being worse off for it. Luck, as defined in this study, exists in the NFL and has a direct impact on the biggest factor, wins in the current year, in deciding whether a coach is retained or fired.

References

- 2007 New England Patriots. (n.d.). *Pro-football-reference.com*. Retrieved from <http://www.pro-football-reference.com/teams/nwe/2007.htm>
- 2008 Detroit Lions. (n.d.). *Pro-football-reference.com*. Retrieved from <http://www.pro-football-reference.com/teams/det/2008.htm>
- About Brian Burke. (2015). *AdvancedFootballAnalytics.com*. Retrieved from <http://www.advancedfootballanalytics.com/index.php/home/about/about-brian-burke>
- Adler, S. E., Berry, M. J., Doherty, D. (2012). Pushing “reset”: The conditional effects of coaching replacements on college football performance. *Social Science Quarterly*, 1-28.
- Anderson, S. (n.d.). Logistic Regression. Retrieved from <http://schatz.sju.edu/multivar/guide/logistic.pdf>
- Arbesman, S. (2012, Nov 16). Luck and skill untangled: The science of success. *Wired.com*. Retrieved from <http://www.wired.com/2012/11/luck-and-skill-untangled-qa-with-michael-mauboussin/>
- Barnwell, B. (2013, July 3). The NFL’s numbers game. *Grantland.com*. Retrieved from <http://grantland.com/features/bill-barnwell-breaks-2012-numbers-get-feel-2013/>
- Barnwell, B. (2013, Aug 23). Thank you for coaching. *Grantland.com*. Retrieved from <http://grantland.com/features/bill-barnwell-scenarios-coaches-fail-most/>
- Barnwell, B. (2014, May 30). Numbers rarely lie. *Grantland.com*. Retrieved from <http://grantland.com/features/nfl-stats-predicting-success/>
- Berri, D. J., & Schmidt, M. B. (2006). On the road with the National Basketball Association’s superstar externality. *Journal of Sports Economics*, 7, 347-358.
- Bill Belichick. (n.d.). *Pro-football-reference.com*. Retrieved from <http://www.pro-football-reference.com/coaches/BeliBi0.htm>
- Bishop, G. (2009). Ravens rookie is young, but he’s the same old Flacco. *NYTimes.com*. Retrieved from http://www.nytimes.com/2009/01/12/sports/football/12ravens.html?_r=0
- Brain, D., and Webb, G. I. 1999. “On the Effect of Data Set Size on Bias and Variance in Classification Learning,” in In Proceedings of the Fourth Australian Knowledge Acquisition Workshop, University of New South Wales, pp. 117–128.

- Breer, A. (2013). Who's really in charge? Power structures vary across NFC East. *NFL.com*. Retrieved from <http://www.nfl.com/news/story/0ap1000000210494/printable/whos-ireallyi-in-charge-power-structures-vary-across-nfc-east>
- Burke, B. (2007, Jul 31). Luck and NFL outcomes. *AdvancedFootballAnalytics.com*. Retrieved from <http://archive.advancedfootballanalytics.com/2007/08/luck-and-nfl-outcomes-1.html>
- Caro, C., & Machtmes, R. (2013). Testing the utility of the pythagorean expectation formula on division one college football: an examination and comparison to the Morey model. *Journal of Business & Economics Research*, 11(12), 537-542.
- Dawson, P, Dobson, S & Gerrard B. (2000). Estimating Coaching Efficiency in Professional Team Sports: Evidence from English association football. *Scottish Journal of Political Economy*. 47 (4), 399-421.
- Deceuster, D., & Kishor, K. (2012, Mar 13). How does parity compare among major U.S. sports leagues? *Forbes.com*. Retrieved from <http://www.forbes.com/sites/quora/2012/03/13/how-does-parity-compare-among-the-major-u-s-sports-nfl-nba-mlb-nhl-college-football-college-basketball/>
- FAQ: What are pseudo R-squareds? *UCLA: Statistical Consulting Group*. Retrieved from http://www.ats.ucla.edu/stat/mult_pkg/faq/general/Psuedo_RSquareds.htm
- Frederick, B. (2013, Apr 1). Fans must understand that college sports is big business. *USNews.com*. Retrieved from <http://www.usnews.com/debate-club/should-ncaa-athletes-be-paid/fans-must-understand-that-college-sports-is-big-business>
- Hadley, L., Poitras, M., Ruggiero, J., & Knowles, S. (2000). Performance evaluation of National Football League teams. *Managerial and Decision Economics*, 21, 63-70.
- Hales, S., & Johnson, J. (2014). Luck attributions and cognitive bias. *Metaphilosophy* 45, 509-528.
- Holmes, P. (2010). Win or go home: Why college football coaches get fired. *Journal of Sports Economics*.
- Kahn, L.M. (2006). Race, performance, pay, and retention among national basketball association head coaches. *Journal of Sports Economics*. 7-(2): 119 – 149.
- Luo, V. (2014). Relieving and readjusting Pythagoras: Improving the Pythagorean formula. Retrieved from http://web.williams.edu/Mathematics/sjmillier/public_html/math/papers/st/VictorL

uo.pdf

- Mason, B. (2014). Why NBA Coaches get fired: You don't know what you don't know, you know? Retrieved from <http://triangleoffense.com/features/why-nba-coaches-get-fired-you-dont-know-what-you-dont-know-you-know/>
- Maxcy, J. (2013). Efficiency and Managerial Performance in FBS College Football: To the Employment and Succession Decisions, which Matters More, Coaching or Recruiting? *Journal of Sport Economics*.14 (4) 368-388
- Meiselman, M. (2012). Win Probability Forfeited 2012. Retrieved from http://webcache.googleusercontent.com/search?q=cache:au_ZoeJY2zIJ:archive.advancedfootballanalytics.com/2013/03/win-probability-forfeited-2012.html+&cd=1&hl=en&ct=clnk&gl=us
- Mihoces, G. (2015). Recycled NFL coaches often find success. *USAToday.com*. Retrieved from <http://www.usatoday.com/story/sports/nfl/2015/01/20/nfl-recycled-coaches-often-find-second-chance-success-in-the-nfl/22071963/>
- Moy, D. (2006). Regression planes to improve the Pythagorean percentage: A regression model using common baseball statistics to project offensive and defensive efficiency. Retrieved from https://www.stat.berkeley.edu/~aldous/157/Old_Projects/moy.pdf
- O'Halloran, S. (n.d.). Lecture 10: Logistical Regression II – Multinomial Data. Retrieved from http://www.columbia.edu/~so33/SusDev/Lecture_10.pdf
- Orzel, C. (2011). Statistical Significance is an Arbitrary Convention. Retrieved from <http://scienceblogs.com/principles/2011/09/12/statistical-significance-is-an/>
- Porter, P. K., & Scully, G. W. (1982). Measuring managerial efficiency: The case of baseball. *Southern Economic Journal*, 48, 642-650.
- Rudy, K. (2014, Jan 17). A statistical look at how turnovers impacted the NFL season. *Minitab.com*. Retrieved from <http://blog.minitab.com/blog/the-statistics-game/a-statistical-look-at-how-turnovers-impacted-the-nfl-season>
- Scully, G. W. (1994). Managerial efficiency and survivability in professional team sports. *Managerial and Decision Economics*, 15, 403-411.
- Swanson, B. (2015, Mar 6). Broncos hire Mitch Tanney as director of football analytics. *DenverBroncos.com*. Retrieved from <http://www.denverbroncos.com/news-and-blogs/article-1/Broncos-hire-Mitch-Tanney-as-director-of-football-analytics/f869fb7d-2692-4c6a-bf99-94d642b93138>

- Vint, P. (2011, Jun 15). Pythagorean expectation and the myth of luck, or how I learned to stop worrying and accept Northwestern. *Blackheartgoldpants.com*. Retrieved from <http://www.blackheartgoldpants.com/2011/6/15/2224697/pythagorean-expectation-and-the-myth-of-luck-or-how-i-learned-to-stop>
- Vogel, B. (2012, Aug 20). Nebraska and Football's Pythagorean Theorem. *HailVarsity.com*. Retrieved from <http://hailvarsity.com/news/football-news/nebraska-and-footballs-pythagorean-theorem/2012/08/>
- Vrentas, J. (2015, Jan 25). Football's focus group. *SI.com*. Retrieved from <http://mmqb.si.com/2015/01/25/pro-football-focus-nfl-neil-hornsby-cris-collinsworth-analytics/>
- Williams, C. (2015, Jul 28). Power struggles becoming the norm for successful NFL franchises. *AthlonSports.com*. Retrieved from <http://athlonsports.com/magazines/power-struggles-becoming-norm-successful-nfl-franchises>
- Wuensch, K. (2015). Binary Logistic Regression with SPSS. Retrieved from <http://core.ecu.edu/psyc/wuenschk/mv/multreg/logistic-spss.pdf>