Draft version November 27, 2018
Typeset using LaTeX **preprint** style in AASTeX61

A NEW EXPECTED POINTS MODEL FOR FOOTBALL DRIVE DATA

Andrew C Guyader, Jeffrey Sklar, and Randy Scovil

ABSTRACT

An expected points model is derived and validated using National Football League (NFL) drive data. The study utilizes NFL regular season games from 2000 through 2016 - over 4,500 games and 100,000 drives - in a dataset spanning 17 complete seasons. Eight years of NFL drive data are used to create the expected points model. The model develops drive expected points as a function of starting field position through an average next points calculation with starting field positions are grouped into five yard increments. The remaining nine years of data are used to evaluate the expected points model through an expected points added (EPA) calculation applied to individual games. This formulation allows for analysis of team performance at both the game and season level. Team performance at the game level is calculated by averaging the EPA values over the entire game (GameEPAavg). The GameEPAavg is analyzed in relation to the point differential of the game. Regression models reveal a strong association between the GameEPAavg and the probability of winning the game. The SeasonEPAavg is the average GameEPAavg for a team across all sixteen games in a full season. SeasonEPAavg is calculated for each team and is shown to be positively associated with the probability of having a winning season - 9 or more wins. Also, EPA from the first half of games are calculated. The halftime EPA values are presented in association with the probability of winning the game.

1. INTRODUCTION

The game of football is flooded with emotion. Fans, players, coaches, executives and owners each have high stakes in a sport where extreme athleticism, speed and power create action seen nowhere else in professional sports. Football has a growing, world-wide audience, where large crowds gather to see elite skill, technique, focus and raw grit. The business of football is *big* and getting bigger. Revenue for the NFL exceeded $13 billion in 2015, up 50 percent from 2010 (Belzer 2016). Additionally, the NFL continues to market its product internationally. In 2016, four regular season games were played in two foreign countries.

For the many entities emotionally and financially invested in game outcomes, there exists a need for useful and accurate tools to help with the reflection, interpretation and tracking of on-field football performance. Football statistics exist at several levels - team, unit, drive, play and player - but the interpretation and application of those statistics in a meaningful fashion has much room for improvement. Extensive data are collected in football games from player tracking devices, play-by-play data and ball tracking data chips. Also, a wide range of bulk statistics exist at the player, unit and team level over quarter, half and complete games. The array of data produced in football games is growing every year.

This paper examines football game data at the drive level. Recorded data for each individual drive includes team, starting field position, clock time, yards covered and drive result. The examination of the game at this granular and chronological level is more revealing than end of game bulk statistics.The drive level analysis interprets yards gained within the context of the starting field position of the drive and unifies the evaluation of all drives on a single success continuum.

2. EXPECTED POINTS

The application of an expected points model to play and drive data yields a methodology to evaluate the effectiveness of on-field performance. The play data will track actual points scored and yardage covered per the down and distance situation. Expected points analysis gives insights beyond the scoreboard and bulk statistics and assists in the quantification of performance on non-scoring drives in a football game.

Pioneering work developing expected points models traces back to a paper by Virgil Carter and Robert Machol (Carter and Machol 1971). In that paper, an expected points model was produced using 1^{st} and 10 situations from NFL play-by-play data from the first 56 games of the 1969 NFL season. The next points calculation was performed using ten field position buckets each with a span of 10 yards. The data set included all 1^{st} and 10 situations including at the beginning of drives and when 1^{st} downs occur within the possession or drive of the football.

The foundational element of the expected points model by Carter and Machol is a next points determination for every drive in a football game. The expected points value was calculated as $E(X) = \sum X_i P(X_i)$. The number of outcomes (X_i) are 103 with four as points scored (touchdown, field goal, safety, turnover returned for touchdown) and an additional 99, one for each yard line of the field where the drive could end. Cabot, et al. further extended the work by using down, yards to go and yardline to derive expected point values (Cabot 1983).

The concept of expected points was further extended in the book *The Hidden Game of Football* (Carroll, et. al. 1988). This book helped establish the public interest in the examination of football statistics beyond bulk game analysis such as per game calculations of yards, turnovers, time of possession, etc. The authors used play-by-play data to reevaluate stats and rethink which ones are important for winning football games. The authors performed the Carter and Machol analysis using 1986 play data and similar results were obtained.

The website Footballoutsiders.com calculated expected points for 1^{st} and 10 situations (FootballOutsiders 2003). David Romer created expected point values for 1^{st} and 10 situations to show that teams punt and kick field goals more frequently than they should (Romer 2006). Brian Burke calculated several play-by-play expected points models at advancedfootballanalytics.com (Burke 2012). Most of the work in expected points, net expected points or expected points added has been formulated at the play-by-play level. The situations encountered in a game featuring down, distance and field position.

An example of the drive data collected, managed and analyzed is shown in Figure 1. Over 100,000 lines of drive data were captured from public-facing drive charts and game logs on several sources including espn.com, cbssports.com, foxsports.com and armchair.com charts and game logs. Data management and quality control was performed through several efforts including visual inspection, score comparisons and clock time tracking. The accuracy of the data set is extremely high across the 17 years of game drive data.

ID	Q5GID	Team	QTR	SFP	TimeStart	HowEnded	Plays	Yards	Points	NextPts	EP	EPAdd
2009010101	20090101	PIT	1	58	15:00	Punt	3	2		7	1.843	0.190
2009010102	20090101	TEN	1	98	13:16	Punt	3	2		-7	-0.741	-2.071
2009010103	20090101	PIT	1	43	11:24	Punt	5	2		7	2.812	-2.652
2009010104	20090101	TEN	1	89	8:20	Missed FG	5	70		-7	-0.160	-0.714
2009010105	20090101	PIT	1	73	6:44	Punt	3	-6		7	0.874	-1.683
2009010106	20090101	TEN	1	74	4:49	Interception	6	30		-7	0.809	-1.296
2009010107	20090101	PIT	1	79	1:38	Interception	3	3		7	0.486	-3.234
2009010108	20090101	TEN	1	44	0:06	Punt	5	0		-7	2.747	-2.200
2009010109	20090101	PIT	2	95	13:04	Punt	8	33		7	-0.547	-0.327
2009010110	20090101	TEN	2	73	7:14	Missed FG	12	60		-7	0.874	-1.360
2009010111	20090101	PIT	2	79	2:14	Touchdown	5	79	7	7	0.486	6.514
2009010112	20090101	TEN	2	71	1:22	Touchdown	3	71	7	7	1.003	6.449
2009010113	20090101	PIT	2	73	0:48	Interception	6	18	0	0	0.874	0.000
2009010114	20090101	TEN	3	79	15:00	Fumble	6	38		3	0.486	-2.652
2009010115	20090101	PIT	3	54	11:57	Punt	4	9		-3	2.101	-1.554
2009010116	20090101	TEN	3	95	9:53	Punt	3	-1		3	-0.547	-1.554
2009010117	20090101	PIT	3	54	8:01	Punt	6	11		-3	2.101	-1.877
2009010118	20090101	TEN	3	90	5:08	Punt	6	19		3	-0.224	0.125
2009010119	20090101	PIT	3	85	1:43	Punt	3	8		-3	0.099	-1.619
2009010120	20090101	TEN	4	63	14:51	Field Goal	7	36	3	3	1.520	1.480
2009010121	20090101	PIT	4	70	11:03	Field Goal	12	56	3	3	1.068	2.449
2009010122	20090101	TEN	4	82	2:57	Punt	4	12		0	0.293	-2.394
2009010123	20090101	PIT	4	58	1:50	Fumble	4	54		0	1.843	-1.231
2009010124	20090101	TEN	4	96	0:51	End Half	3	0	0	0	-0.612	0.000
2009010125	20090101	PIT	5	78	15:00	Field Goal	10	63	3	3	0.551	2.449

Figure 1. Drive data for the performance metric including elements of database storage.

2.1. *Drive Endings*

The possible drive results or outcomes at the conclusion of any possession are the following: touchdown (TD), field goal (FG), missed FG, punt, safety, turnover (by fumble or interception), downs (a failed 4^{th} down conversion), end-of-half and end-of-game. Although rare, several possessions may occur within a single play. For instance, a fumble returned for a touchdown would be two possessions - one for each team with a zero play drive for the scoring team. In all cases, the possession of the football is determined by the officials. Here forward, each possession is termed a *drive*.

A summary for drive endings in NFL games is shown in Figure 2. This analysis reveals that punts occur on almost 40% of possessions. Point scoring possessions - touchdowns, field goals and safeties - occur on approximately one out of every three drives. Drives end with a turnover, fumble or interception on just under 15% of possessions.

Drive endings by starting field position are shown in Figure 3. The horizontal axis represents the starting field position (SFP) in 5 yard increments. SFP is the yardage to the endzone at the start of the drive. The bin labelled *0* covers SFP from 1 to 4 yards, the bin labelled *5* covers 5 to 9, etc, the bin labelled *90* covers 90 to 94, the bin labelled *95* covers 95 to 99 yards. The drive ending options are color coded showing percentage of occurrence in each bin. The large count in the 80 bin reflects the large number of possessions that start after a touchback.[1]

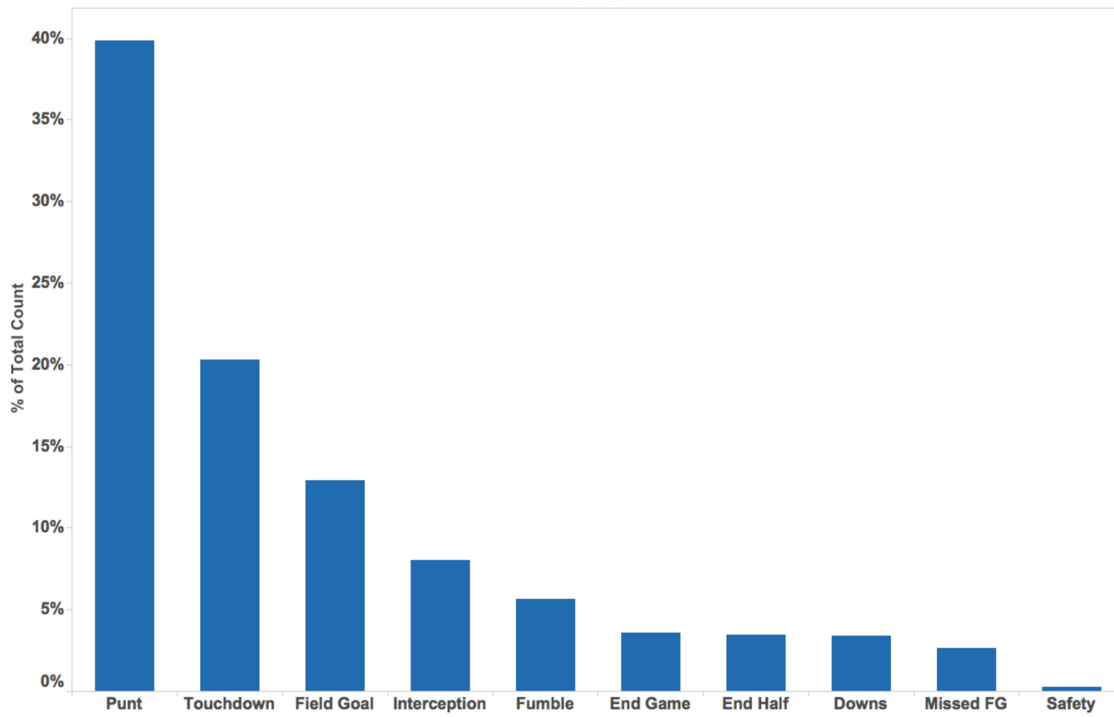

Figure 2. Drive results.

2.2. *Next Points Calculation*

Drive data are arranged chronologically and tracked by game. Possessions that ended with points: touchdowns, field goals and safeties were identified. The touchdowns are assigned a value of *7*, field goals a *3* and safety is a *-2*.[2] Also, drives that end the half or game are assigned a value of *0*.

Working backwards chronologically, the value from the points scored is multiplied by *-1* on each proceeding drive, stopping when a drive is encountered with a point value already assigned. This is shown in Figure 1 where the column titled *NextPts* tracks the calculation. Figure 4 shows the next points mean values as a function of SFP.

[1] Before the 2016 season, the touchback line was changed from the 20-yard line to the 25-yard line.

[2] For drives resulting in a touchdown, the data set does not track what happened on the points after touchdown play. The analysis assumed that a touchdown resulted in 7 points when in fact the result could be 6, 7, 8 or in the rare case the team on defense can return a point after attempt for two points meaning that the net points was only 4.

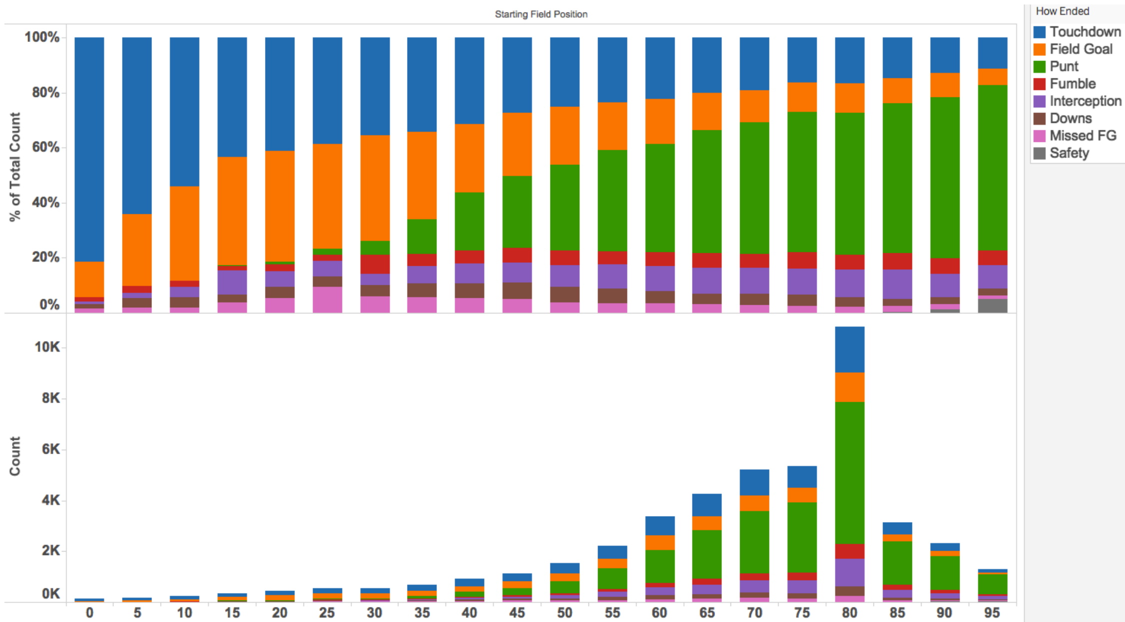

Figure 3. Drive results by starting field position excluding end of half or game.

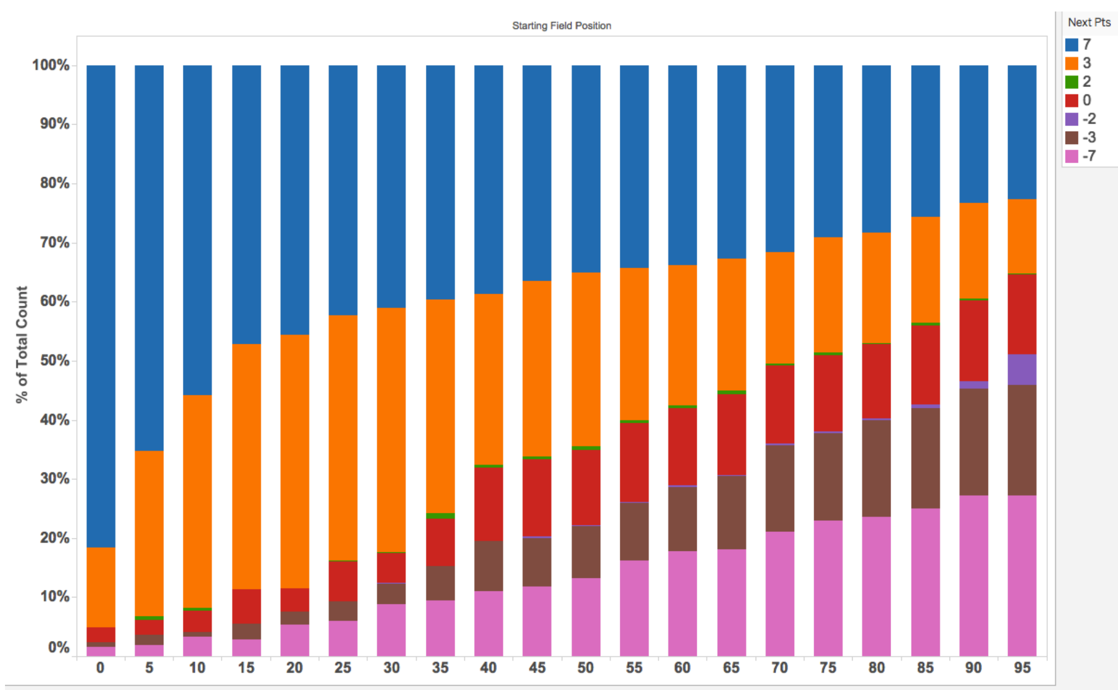

Figure 4. Next points visualization for odd years of NFL data from 2000-2016.

2.3. *Expected Points Regression and Expected Points Added*

A linear regression on the average next points calculation is shown in Figure 5. The expected points (EP) equation using eight years of NFL data (the odd years from 2000-2016) is:

$$EP(i) = -0.0646 * SFP(i) + 5.590 \tag{1}$$

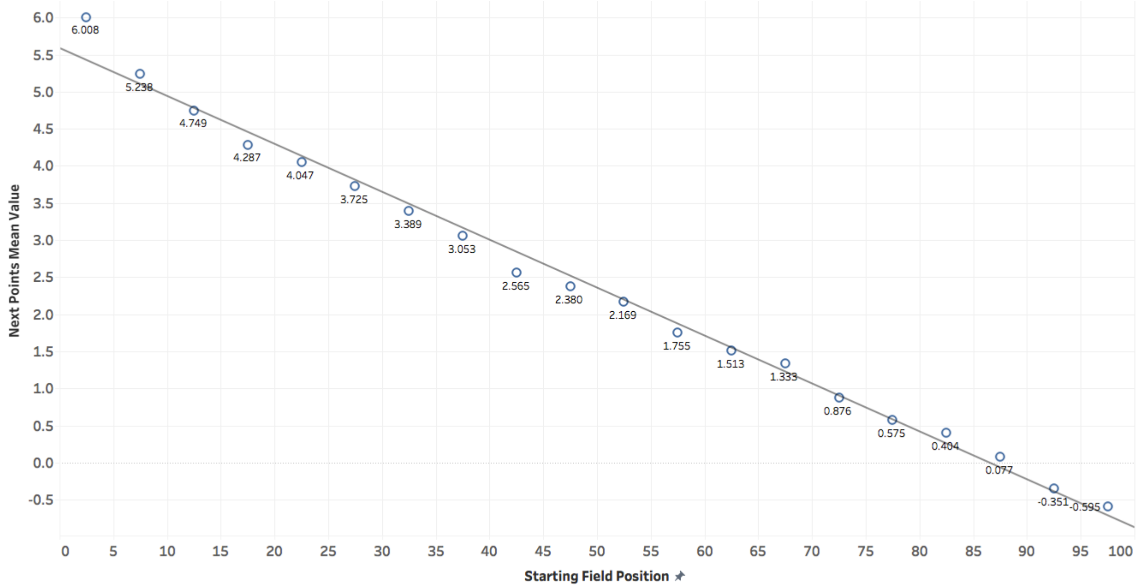

Figure 5. Next points regression for odd years of NFL data from 2000-2016.

The application of the EP model to individual lines of data within the context of a game further generates the need to track the team in possession of the football. The expected points derived thus far is from the perspective of the team in possession of the football and this perspective will continue throughout.

Expected Points Added (EPA) quantifies the performance of a drive by taking into account the drive result, points scored and starting field position values. The index i references the team in possession of the ball on drive i. If no points are scored on a drive, the calculation for EPA on drive i is:

$$EPA(i) = -[EP(i+1) + EP(i)] \tag{2}$$

If Points are scored on drive i by the offense then the EPA on drive i is calculated as

$$EPA(i) = Points(i) - EP(i) \tag{3}$$

Equation 3 is applied only when points are scored by the offense from a scrimmage play on a drive where possession was initiated after a punt or turnover. The possible values for $Points(i)$ are 7 for a touchdown, 3 for a field goal and -2 for a safety. Other point scoring situations are presented in Appendix A.

2.4. *Game Expected Points Added Average*

The average game expected points is the average of all expected points added values in a single game. The GameEPAavg is calculated as:

$$GameEPAavg_{TEAM_A} = \frac{\sum_{i=1}^{n_a} EPA_{TEAM_A}(i) - \sum_{i=1}^{n_b} EPA_{TEAM_B}(i)}{n_a + n_b} \tag{4}$$

where n_a is the number of possessions for $TEAM_A$ and n_b is the number of possessions for $TEAM_B$. It follows then that the GameEPAavg value for each team is related by

$$GameEPAavg_{TEAM_B} = -GameEPAavg_{TEAM_A} \tag{5}$$

2.5. *Season Expected Points Added Average*

The entire 16 games in a season are averaged together to calculate a Season Expected Points Added value (SeasonEPAavg). The SeasonEPAavg calculation is

$$SeasonEPAavg = \sum_{i=1}^{16} GameEPAavg(i)/16 \tag{6}$$

The SeasonEPAavg values are analyzed in relation to season win totals in Section 3.2.

3. ANALYSIS

3.1. *Association between GameEPAavg and Winning Point Differential*

Analyses were conducted to examine the association between the GameEPAavg and the point differential (winning team score minus losing team score). This analysis was performed on nine complete seasons of 256 games per season (the even number years from 2000-2016). It was of particular interest to determine whether the GameEPAavg is a significant predictor of the point differential. Figure 6 displays the scatterplot relating GameEPAavg and point differential with a linear regression model.

The analysis revealed a significant positive association between the GameEPAavg and winning team point differential. The regression coefficient and standard error is shown in Table 1. Each additional point increase in GameEPAavg is associated with a 17.4 point increase in winning point differential, on average. For the regression model, the R^2_{adj} value is 0.883.

	Coefficient	Standard Error
GameEPAavg	17.414*	0.132

Table 1. Linear regression results for GameEPAavg (*p-value < 0.001).

3.2. *Association between SeasonEPAavg and Odds of Winning More than 8 Games*

The season win total is plotted against the SeasonEPAavg value for the even-numbered years between 2000 - 2016. This yields 288 data points, representing 32 teams over 9 seasons. The data and regression are shown in Figure 7. The regression shows a strong association between SeasonEPAavg and win total. The linear regression results are shown in Table 2. The R^2 value is 0.826. A one unit increase in the SeasonEPAavg value relates to a season win total increase of 8.35.

Further analysis also reveals that there is a significant positive association between SeasonEPAavg and the odds of having a winning season. Each additional 0.1 point increase in SeasonEPAavg is associated with between a 167% and 413% increase in the odds of having a winning season (winning 9 or more games). The results of the binary logistic regression are shown in Table 3 with the 95% Confidence Interval (CI) shown in the far right column.

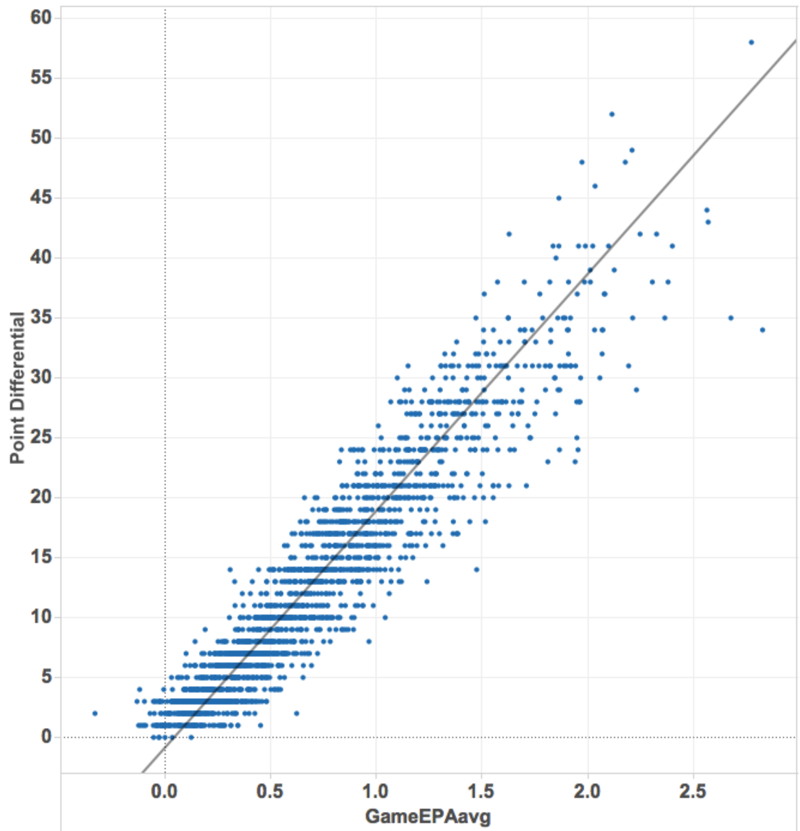

Figure 6. Point differential vs GameEPAavg

	Coefficient	Standard Error
SeasonEPA	8.349*	0.230

Table 2. Linear regression results for SeasonEPAavg (*p-value < 0.001)

	Coefficient	Standard Error	Estimated Odds Ratio (OR)	95% CI for True OR
SeasonEPA	13.09*	1.66	3.70	(2.67, 5.13)

Table 3. Binary logistic regression results for SeasonEPA and season wins > 8 games (*p-value < 0.001).

3.3. *Association between Halftime EPA Average and Winning the Game*

The halftime expected points added average (HalfEPAavg) is Eqn 4 but only includes drives from the first half of games. The distribution of the HalfEPAavg values for the eventual winning team is shown in Figure 8.

For the entire NFL, there is a significant positive association between HalfEPAavg and the odds (and hence probability) of winning a game. The logistic regression coefficient and standard error are shown in Table 4. Each additional 0.1 point increase in HalfEPAavg is associated with between a 93.9% and 153.7% increase in the odds of winning the game.

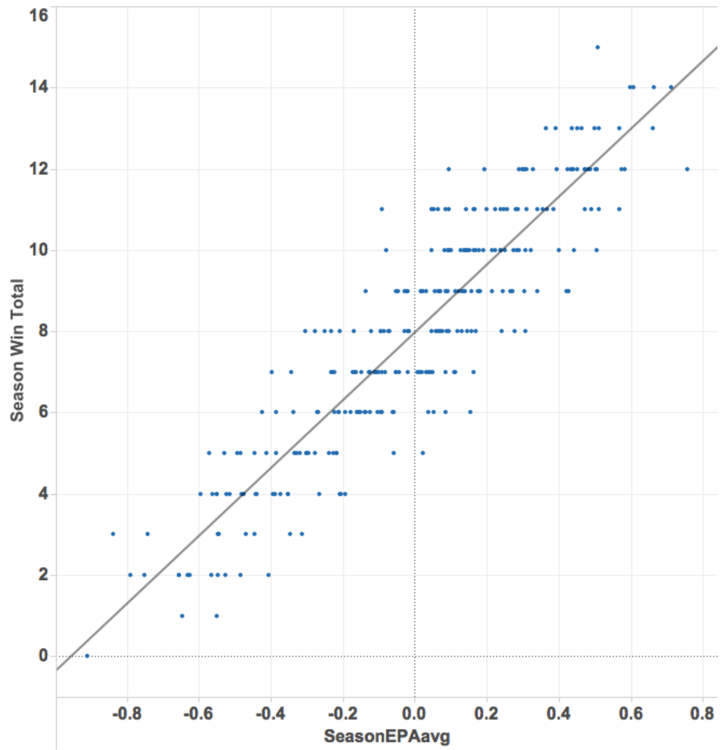

Figure 7. Season win total vs SeasonEPAavg

	Coefficient	Standard Error	Estimated Odds Ratio (OR)	95% CI for True OR
HalfEPAavg	2.568*	0.116	13.0405	(10.385,16.375)

Table 4. Binary logistic regression results for Association between Halftime EPA and winning the game (entire NFL). (*p-value < 0.001).

The halftime analysis was also performed for each team individually. The estimated odds ratio corresponding to a one point increase in HalfEPAavg, as well as the corresponding 95 percent confidence interval for the true odds ratio are shown in Table 5. For each team, there is a significant positive association between HalfEPAavg and the odds of winning the game. The limits of the 95 percent confidence intervals for the true odds ratios associated with a point increase in HalfEPAavg are above 1.0 for all teams. Note the wide variability in the confidence interval limits for the teams suggesting varying levels of association between the HalfEPAavg and the outcome of the game.

A closer examination of Table 5 shows that the Bills, Patriots, Steelers and Cardinals have the highest odds ratios. For these teams playing better in the first half has a strong effect on their ability to win the game compared to other NFL teams. The teams at the other extreme are the Chargers, 49ers, Chiefs and Jaguars. For these teams, playing better in the first half is less important in relation to their ability to win the game.

4. CONCLUSIONS

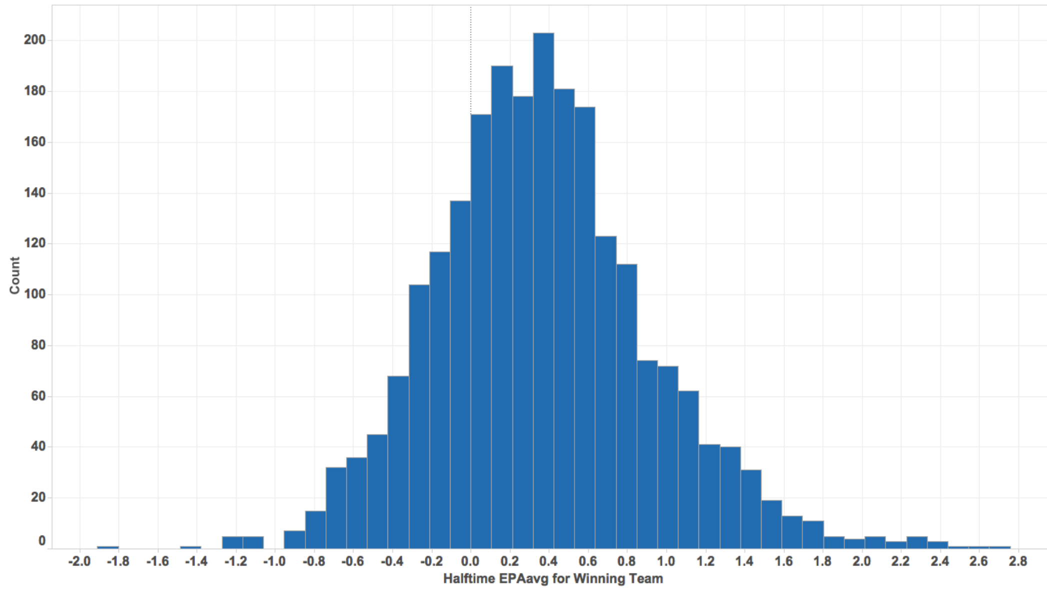

Figure 8. Winning team HalfEPAavg from even year data

The expected points model and expected points added formulation is a useful aide to evaluate and track on-field football performance. The expected points analysis aides to separate human emotions from necessary objective analysis in this highly emotional and scoreboard-driven industry.

The expected points model developed in this study shows accurate analysis of team performance throughout a football game. The expected points added calculation for each strongly associates with the final score point differential. Each additional point increase in GameEPAavg is associated, on average, with a 17.4 point increase in winning point differential. The linear regression model yields an R^2_{adj} value of 0.883.

A season analysis over 16 games reveals that a one point increase in the SeasonEPAavg value relates to a season win total increase of 8.35. The linear regression results have an R^2 value is 0.826. This season-level calculation reveals a strong association of the SeasonEPAavg to the odds of having a winning season (winning more than 8 games). Each additional 0.1 point increase in SeasonEPAavg is associated with between a 167% and 413% increase in the odds of having a winning season.

A first half only analysis (HalfEPAavg) shows a strong association with the odds of winning the game. Each additional 0.1 point increase in HalfEPAavg is associated with between a 93.9% and 153.7% increase in the odds of winning the game. This analysis is performed for both all teams together and for individual teams, revealing different odds ratios for each team.

Team	Estimated Odds Ratio (OR)	95% CI for True OR
Arizona Cardinals	25.884	(8.273, 80.980)
Atlanta Falcons	13.784	(5.542, 34.284)
Baltimore Ravens	13.770	(5.098, 37.193)
Buffalo Bills	35.683	(10.289, 123.755)
Carolina Panthers	17.030	(6.433, 45.083)
Chicago Bears	21.889	(7.497, 63.912)
Cincinnati Bengals	23.319	(8.128, 66.900)
Cleveland Browns	7.505	(3.061, 18.399)
Dallas Cowboys	13.594	(5.082, 36.368)
Denver Broncos	12.995	(5.516, 30.612)
Detroit Lions	8.188	(3.451, 19.430)
Green Bay Packers	18.897	(6.880, 51.904)
Houston Texans	16.314	(6.007, 44.311)
Indianapolis Colts	7.754	(3.786, 15.879)
Jacksonville Jaguars	7.492	(3.266, 17.184)
Kansas City Chiefs	7.230	(3.428, 15.250)
Los Angeles Chargers	5.083	(2.714, 9.520)
Miami Dolphins	14.611	(5.466, 39.058)
Minnesota Vikings	14.174	(5.621, 35.739)
New England Patriots	33.718	(9.848, 115.440)
New Orleans Saints	24.337	(8.572, 69.099)
New York Giants	10.102	(4.286, 23.808)
New York Jets	18.149	(6.451, 51.060)
Oakland Raiders	11.274	(4.961, 25.620)
Philadelphia Eagles	18.396	(6.578, 51.060)
Pittsburgh Steelers	31.361	(9.315, 105.587)
Los Angeles Rams	10.583	(4.168, 26.873)
San Francisco Forty-Niners	7.200	(3.408, 15.215)
Seattle Seahawks	19.068	(7.087, 51.215)
Tampa Bay Buccaneers	19.580	(7.002, 54.755)
Tennessee Titans	7.385	(3.501, 15.578)
Washington Redskins	11.005	(4.233, 28.612)

Table 5. Estimated Odds Ratios and 95% Confidence Intervals for True Odds Ratio corresponding to a 1 unit increase in the HalfEPAavg for all 32 NFL teams

Appendix A

If points are scored by the offense after the team receives the ball on a kickoff

$$EPA(i) = Points(i) - EP_{78} \tag{7}$$

where EP_{78} is the expected points values for drives starting 78 yards from the end zone. 78 yards is the average starting field position calculated from the dataset across all drives. In this fashion, the kickoff return performance is accurately managed in the EPA calculation. If a kickoff is run back for a TD, that kickoff will be drive i in the following calculation:

$$EPA(i) = 7 - EP_{78} \tag{8}$$

A kickoff occurring in the game sets a special calculation when the ensuing drive does not end with points scored. Working from the concept that the averaging starting field position is 78 yards from the end zone, the EPA from a non-scoring drive after a kickoff is calculated using the following expression:

$$EPA(i) = -[EP_{78} + EP(i+1)] \tag{9}$$

If the offense turns the ball over and the defense returns that turnover for a TD or if a possession ends with a punt by the team on offense and that kicking play results in a TD for the team set to receive the punt, that offensive possession will be drive i in the following:

$$EPA(i) = -7 - EP(i) \tag{10}$$

These are two examples of zero play touchdowns where the line of data for the zero play TD has an EPA value of zero. From Eqn 10, $EPA(i+1) = 0$.

Any drive with the result end-of-half or end-of-game will have an EPA value equal to zero.

For completeness, the following scenarios are addressed:

- Team A performs a kickoff, Team B possesses the kicked football [drive (i)] but fumbles during the return and Team A recovers [drive $(i+1)$]. This situation would use Equation 9.
- Team A performs a kickoff, Team B possesses the football [drive (i)] but fumbles and Team A recovers [drive $(i+1)$] and runs it back for a touchdown, then $EPA(i) = -7 - EP_{78}$ and $EPA(i+1) = 0$.
- On drive (i), Team A fumbles or punts on offense, Team B possesses the football [drive $(i+1)$] and as the same play continues Team B fumbles and team A recovers. Team A is now back in possession of the football and the play ends. On the next play, a play is run from scrimmage thereby defining drive $(i+2)$. EPAs are calculated as follows: $EPA(i) = -[EP(i+3) + EP(i)]$ if no points are scored on drive $i+2$ or $EPA(i) = Points(i+2) - EP(i)$ if points are scored on drive $i+2$. In both cases, $EPA(i+1)$ and $EPA(i+2)$ equal zero.
- On drive (i), Team A fumbles or punts on offense, Team B possesses the football [drive $(i+1)$] and as the same play continues, Team B fumbles and Team A recovers. Team A is now back in possession of the football and still on the same play, scores a touchdown [drive $(i+2)$]. EPAs are calculated as follows: $EPA(i) = 7 - EP(i)$. $EPA(i+1)$ and $EPA(i+2)$ both equal zero.

REFERENCES

Alamar, Benjamin. (2013) *Sports Analytics*, Colombia University Press, New York.

Bassett, G.W. (1997). *Robust Sports Ratings Based on Least Absolute Errors*, The American Statistician, 39, 152-155.

Belzer, Jason. *Thanks To Roger Goodell, NFL Revenues Projected To Surpass $13 Billion In 2016* Taken from: http://www.forbes.com.

Brimberg, J. and W.J. Hurley. (1997). *The Turnover Puzzle in American Football*, Technical Report - Royal Military College of Canada.

Burke, Brian. (2012). *Expected Points and Expected Points Added Explained*, Taken from: http://www.advancedfootballanalytics.com.

Burke, Brian. (2014). *In-game Win Probabilities*, Taken from: http://www.advancedfootballanalytics.com.

Cabot, Sagarin and Winston. (1983). *????* Operations Research.

Carroll, Bob, Pete Palmer and John Thorn. (1988). *The Hidden Game of Football*, Warner Books, New York.

Carter, Virgil and Robert Machol. (1971). *Technical Note - Operations Research on Football*, Operations Research. 19(2):541-544.

www.footballoutsiders.com.

Goode, B. (1971). *Relevant Variables in Professional Football*, ASA Proceedings of the Social Statistics Section. 83-86.

Glickman, M. E. and H. S. Stern. (1998). *A State-Space Model for National Football League Scores*, Journal of the American Statistical Association. 93:25-35.

Hanafee, Connor. (2013). *Data Analytics in Support of Army Football Offensive Operations*, Honors Thesis United States Military Academy at West Point.

Harville, D. (1977). *The Use of Linear Model Methodology to Rate High School and College Football Teams*, Journal of the American Statistical Association. 72:278-289.

Harville, D. (1980). *Predictions for National Football League Games via LInear Model*, Journal of the American Statistical Association. 75:516-524.

Kaslis, D. and I. Ntzoufras. (2003). *Analysis of Sports Data by using Bivariate Poisson Models*, The Statistician 52(3):381-393.

Keller, J. B. (1994). *A Characterisation of the Poisson Distribution and the Probability of Winning a Game*, The American Statistician. 48(4):294-298.

Michael Lewis (2004) *Moneyball* W.W. Norton and Company Ltd., New York, NY.

Lock, Dennis and Dan Nettleton. (2014) *Using random forests to estimate win probability before each play of an NFL game* Journal of Quantitative Analysis in Sports. Vol.10, Issue 2.

Maki, Kahlan. (2013). *Data Analytics in Support of Army Football Defensive Operations*, Honors Thesis United States Military Academy at West Point.

Palmer, Pete. (1989). *Hidden Game of Football*, .

Pattani, Alok. Expected points and EPA explained. http://www.espn.com/nfl/story/_/id/8379024/nfl-explaining-expected-points-metric 2012

Pelechrinis, Konstantinces and Evangeleos Papalexakis. (2016) *Footballonomics: The Anatonomy of American Football.*

Quenzel, Jared and Paul Shea. (2016). *Predicting the Winner of Tied National League Football Games: Do the Details Matter?*, Journal of Sports Economics. 17, 661-671.

Romer, David. (2006). *Do Firms Maximize? Evidence from Professional Football*, Journal of Politcal Economy. 114:2, 340-365.

Silver, Nathan. (2012). *The Signal and the Noise*, Penguin Group, USA.

Skinner, Brian. (2012). *Scoring Strategies for the Underdog: A General, Quantitative Method for Determining Optimal Sports Strategies*, Journal of Quantitative Analysis in Sports. Vol. 7, Issue 4.

Stern, H. S. (1991). *On the Probability of Winning a Football Game*, The American Statistician. 45:179-183.

Stern, H. S. (1994). *A Brownian Motion Model for the Progress of Sports Scores*, Journal of the American Statistical Association. 89:1128-1134.

Stern, H. S. (1997). *American Football*, Department of Statistics, Iowa State University.