Predicting NFL Game outcomes using Probability models

Bruno Helmeczy

01/01/2021

Executive Summary

This report investigates association patterns between NFL game outcomes & year-to-date team statistics per game. It investigates the research question "How well can NFL game Wins & Losses be predicted based on playing teams' per game statistics ?" & serves to conclude games' expected results, from which unexpected results can further be investigated. From the analysis below, readers can learn competing teams' most significant statistics for explaining game outcome, how these key statistics are associated to derive win probabilities, & how well year-to-date team statistics explain what happened in a game.

For accuracy assessment, probabilities were classified into wins & losses with a 0.5 threshold value. The final Probit-model incorporates Offensive & Defensive Passer Rating differential, Offensive Rushing Yards/Play, Sacks suffered & Offensive Turnovers. It can accurately predict 75.8% of game outcomes in sample (the 2020 season thus far) & tested 69.6% accurate on the 2019 season (though only Offensive & Defensive Passer Rating differentials were found robust). Though artificial neural networks have achieved better accuracy (Kahn, 2003), the presented model rivals expert analysts Purucker (1996) used as baseline comparison, & greatly outperforms ESPN expert analysts Kahn (2003) used for comparison. It also rivals / beats more recent NFL game prediction efforts, with Warner (2010) achieving 64.4% with a Gaussian process model using 10 years' data, Shau (2011) achieving 68.4% with Support Vector Machine (SVM) model using 41 years' data, & Owen & Galle (2014) achieving 68.6%, also with SVM & 5 years' data.

Introduction: This report investigates association pattern with regards to NFL games & looks to answer the research question "How well can NFL game Wins & Losses be predicted based on playing teams' per game statistics ?". I hesitate to call this analysis predictive, as teams' year-to-date statistics are compiled from the games, whose outcomes are the object of prediction. However, this analysis serves to conclude what was the expected result for a game, from which unexpected results can further be investigated. In essence, from this analysis, readers can learn **a.** competing teams' most significant year-to-date statistics which explain a specific games' outcome, **b.** how these statistics are associated to win probabilities, & **c.** how well year-to-date team statistics explain what happened in a game.

American football is a game of territorial gain. It is played with distinct offensive- & defensive units of 11 players on the field. Team A's offense takes on Team B's defence & vice versa. Offenses' objective is to march through a field of 100 yards & reach the opponents end zone, resulting in a Touchdown, worth 7 points. The team with more points at the end of the game wins. Offenses must gain at least 10 yards in 4 attempts, & are rewarded with a 1st down, i.e. another 4 attempts to gain 10 yards. How many yards a team needs to march from how many attempts is communicated as e.g. 1st & 10. This means it is the offenses' 1st of 4 attempts (called downs) & they need to gain 10 yards to receive a 1st down. If team A's offense loses the ball, or cannot gain the yards needed, team B gets the ball from the spot team A got to. Then, team A's defense plays team B's offense, & team B tries to reach the opposite end zone for a touchdown.

Offense: To get the yards needed, offenses can either run (i.e. execute a running play), or pass the ball (i.e. execute a passing play). Every play starts with the the ball snapped to the Quarterback, whom can either hand the ball off to a runningback, or pass it to a receiver.

Running: When the Runningback gets the ball, he runs to gain as many yards as possible, until being tackled to the ground by the defense. The next play starts from where the runningback was tackled.

Passing: Quarterbacks can also keep the ball & throw to their receivers. A receiver catching the ball before it hits the ground is a completed pass, & the receiver can run forward just like runningbacks, until tackled. The next play starts from where the receiver was tackled. If the ball touches the ground before being caught, it is an incomplete pass & the next play starts where the previous did. Also, the gameclock stops between plays after an incomplete pass, another factor offenses consider when deciding to pass or run.

Defense: Many things can go wrong on either play, precisely defenses' job to enforce. The quarterback may be tackled before throwing the ball, called a **Sack**. Both he & the runningback can lose the ball before throwing or being tackled, called a **Fumble**. Then, defensive players can also collect the ball, & the opponent gets to play offense, if successful. The quarterbacks' pass can also be caught by defensive players, an **Interception**, again resulting in the opponent to play offense. Fumbles & Interceptions are together recorded as **Turnovers**. Just as offenses choose if they should pass or run, defenses can focus on either too.

Defending the Run or Pass: A Pass-oriented defense play would assign 3-4 players to sack the quarterback, & 7-8 players to cover receivers. The aim is to enforce an error: The quarterback wants to pass to receivers in good position to catch the ball, but 7-8 players are defending them, so it takes more time for them to find an open position, but he cannot wait long, else the defense will succeed sacking him, resulting in lost yards. This is called "pressuring the quarterback". Alternatively, defenses can focus on running plays by assigning 7-8 players to tackle the Runningback or Quarterback, leaving 3-4 pass defenders.

The Games' tactical nature: Deciding if the offense should run or pass on a play broadly depends on:

1) How many yards the offense needs to gain in the next how many attempts ?
2) How much time is left from the game & what is the current score ?
3) How effective they are at passing or at running the ball ?
4) How good is the opponent at defending against passing or running the ball ?

Each play starts with snapping the ball to the quarterback & ends either with the ball-carrier tackled, running out of bounds, an incomplete pass, or a touchdown. When a play ends, the offense has 40 seconds to decide & start their next play (recall, if the play ended with a ball carrier being tackled, the game clock keeps moving in these 40 seconds). In most cases, this is the head coaches' decision. What is more meaningful, is the quarterbacks' role. The coach communicates the play to execute to the quarterback, whom distributes this information, then the team takes their pre-snap position, & the defensive unit takes their positions, reacting to the offense. At this point the quarterback surveys the defense (judging the defensive formation) & he's free to change the play coaches wanted to execute. If so, he calls a pre-snap audible, screaming code words to his team-mates, whom now know their new assignments. This highlights the Quarterbacks' significance.

Quarterbacks: Every play goes through the quarterback: He touches the ball in every play, & he distributes what play is happening, with the authority to change said play last-second. He is also the player making the most complex decisions during plays & certainly the most influential position to an American Football teams' success. An audible to execute a run, not a pass, due to recognizing a weakness in the defense is equally important to a teams' success, as changing to a different passing play. Thus, it is said by expert analysts that a good quarterback also makes his teammates better, most obviously by putting them in better situations to succeed, leading to better measure values in seemingly unrelated statistics. Quarterbacks are so important, that a separate metric quantifies their performance, the Passer Rating.

Statistical Measures: Measuring how good teams' offensive / defensive units are is an elusive concept & a number of statistics are collected to characterize their success in both running & passing. If running & passing are well characterized however, the same statistics can capture to the same extent offensive & defensive performance. In other words, how many yards a teams' offense gains per running play helps to characterize the teams' offense as much as how many yards a teams' defense allows per running play.

Running Statistics: To characterize a teams' running offense, one may be interested in the **a.** average yards gained per running play, **b.** number of running plays executed, **c.** total rushing yards, or **d.** touchdowns.

Average yards per play establishes expected gains of a running play. If e.g. 5, it means a team will get to 2nd & 5 on average, needing 5 yards in 2 more plays. Since they can expect 5 yards on any running

play, they are in perfect position to try a long-pass play, with high upside without much to lose. Also, opponents' defenses know their success in running & must pay more attention to defending the run, leaving pass defenders 1-on-1 vs receivers, allowing better chances for a long-pass to succeed. **Running plays per game** on the other hand can be thought of as the teams' tactical commitment to running & a proxy for clock management, or teams' time of possession, since running the ball implies the game clock ticking between plays. Frequently against star quarterbacks, opponents' best defensive strategy is for their offense to run the ball, thus keeping said quarterback off the field.

Passing Statistics: To characterize a teams' passing offense, one might be interested in the **a.** average yards gained per passing play, **b.** number of passing plays executed per game, **c.** number of passes completed per game, **d.** touchdowns per game, or **e.** number of times the ball was intercepted per game.

Sacks: Under how much pressure defenses keep quarterbacks is the primary mechanism by which they force mistakes. The most accurately measurable statistic for this are Sacks. Though what is intended to be captured, is how much defenses can alter quarterbacks' decision-making, for every Sack a defense accomplishes, they **Hit** the quarterback few times (tackle him just after throwing), & for every hit, they also **Hurry** him. These are consistent ratios, so Sacks can represent the pressure quarterbacks experience.

Passer Rating: As aforementioned, a separate metric was invented to capture quarterbacks' performance, given their significance in winning American football games. This NFL-specific measure is scaled to be between 0 & 158.3 (the higher, the better), & incorporates the 5 metrics mentioned above. 2 major flaws in the passer rating metric is its' negligence of a. How many Sacks did the passer encounter & b. How many times did he lose the ball via Fumbles. For the specific formula, please see the appendices.

My Data

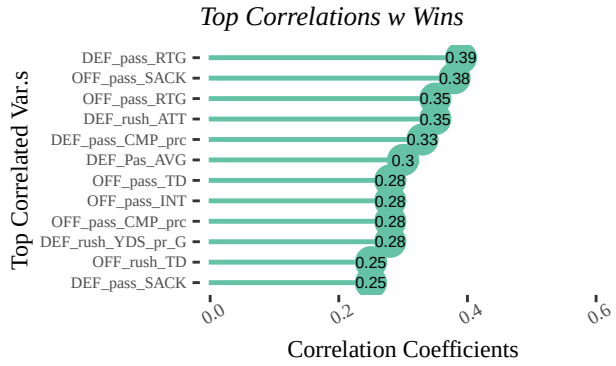
The data for this analysis comprises the 2020 NFL season & was scraped from ESPNs' website. **1st** the 2020 NFL regular seasons' game outcomes before January 1st 2021, & **2nd** the following Year-to-Date Team statistics accumulated over the season were collected (averaged per game unless noted otherwise & collected for offenses & defenses): Passing **a.** Attempts, **b.** Completions, **c.** Completion %, **d.** Yards per Attempt, **e.** Yards, **f.** Touchdowns, **g.** Interceptions, **h.** Sacks, **i.** Passer Rating, & Rushing **j.** Attempts, **k.** Yards per Attempt, **l.** Yards, **m.** Touchdowns, **n.** Fumbles Lost, plus **o.** Turnovers sum Interceptions & Fumbles Lost. For defenses, these are to be interpreted as what opponents' offenses accomplished against them. This resulted in 30 variables collected per team. Finally, Away teams' year-to-date statistics have been divided by Home teams' respectively. Keeping Home- & Away teams (i.e. Opponent & Team), & all game outcomes (the dependent variable) yielded 240 observations & 33 variables. The binary dependent variable takes the value of 1 (Win) if the Away team won, & 0 otherwise.

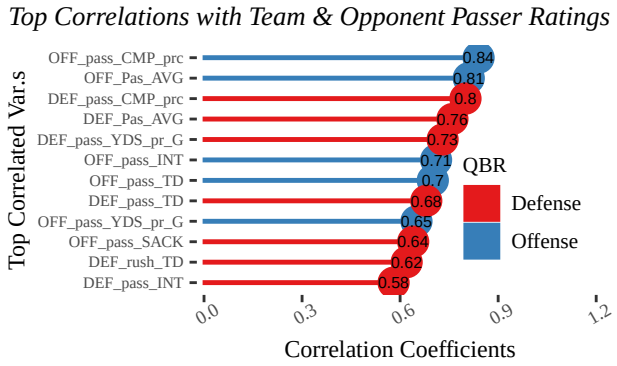
Dataset weaknesses: While all variables were conveniently included in the preceding discussion on American Football, a few additions could be useful for better predictions substantively. Weather conditions in open stadiums can impact play calling significantly: pouring rain causes a muddy field & slippery ball (harder to catch & easier to drop). Injuries are big part of football, & can alter teams' prospects significantly.

	Mean	StDev	Skew	Min	1st IQR	Median	3rd IQR	Max
OFF_pass_RTG	1.01	0.18	0.44	0.63	0.88	1.00	1.13	1.53
OFF_pass_SACK	1.13	0.61	2.11	0.27	0.74	0.99	1.33	4.75
DEF_pass_RTG	1.01	0.12	0.34	0.70	0.94	1.00	1.08	1.46
OFF_rush_AVG	1.01	0.13	0.36	0.68	0.91	1.00	1.10	1.47
OFF_Turnovers	1.37	1.16	1.99	0.16	0.52	1.01	1.78	6.83

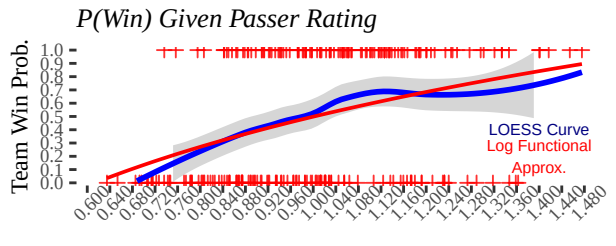
Summary Stats & Histograms: Taking ratios of YTD variables leads to the general interpretation of how much better the Away teams' statistic was vs the Home team. While this seems intuitive substantively (How much better a team played over the other is a frequent explanation for why they have won among analysts), though might right-skews variables. The visual checks (available in the appendices) confirm this.

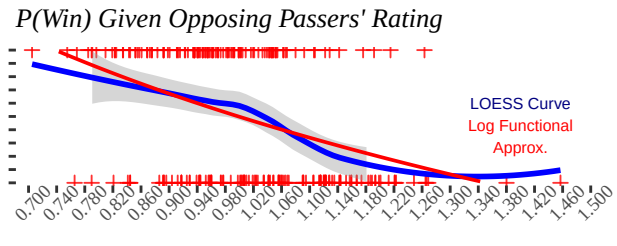
Correlations were used as starting points, in hopes of **a.** finding strongest correlating variables with Wins & **b.** finding variables that strongly correlate with other possible features, enabling to simplify our model, yet cover most available information. Please see the appendices for a Correlogram of all 465 variable pairs. The key highlight from the plots below is the importance of the relative performance in Quarterbacks' Passer Ratings, both how much better a teams' QB plays vs the opponents (OFF_pass_RTG), & how much better do teams defend opponents' QBs vs opposing teams (DEF_pass_RTG). Also, note on the right-side how strongly other variables correlate with Offensive / Defensive Passer Ratings. Passer Rating incorporates Passes, Completions, Yards, Touchdowns & Interceptions, while Completion% & Pass Average are derived from these measures, thus Passer Ratings cover 7 of 8 Passing-related variables, Sacks being the 1 passing statistic not incorporated. Thus, a sensible 1st model would use Offensive & Defensive Passer Rating as explanatory variables. Additionally, Sacks show the 2nd strongest correlation with Wins, though offensive Sacks correlate with Defensive Passer Ratings, while 3 Rushing-related stats correlating with Win probability are Rushing Attempts, Touchdowns & Yards. Rushing attempts may possibly act as proxy for time of possession (recall the game clock continues after all running plays), while Yards & Touchdowns are Rushes' success measures. Note, Rushing Yards/Play combines Yards & Attempts, thus combines the seemingly most important Rushing-related variables.

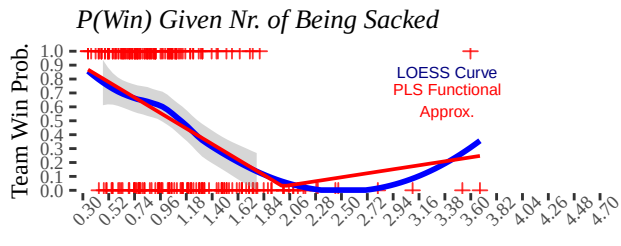


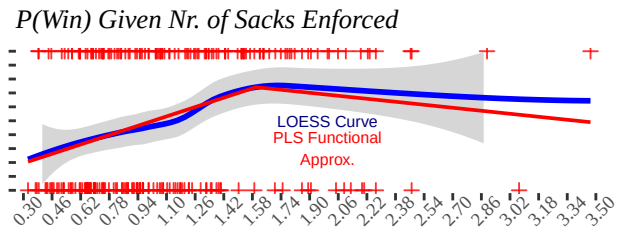


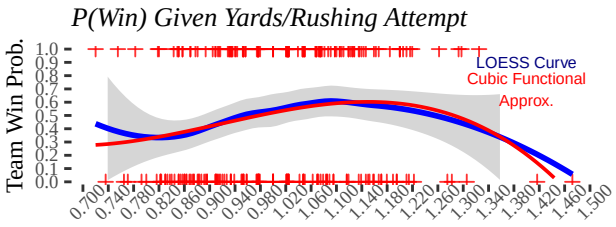
Association Patterns: Please see below the association patterns between these variables & win probability, together with the best-approximating functional forms. As visible from the plots, Conditional Win Probabilities given relative Passer Rating performance stats can be best approximated with a log-functional form, relative Sack stats with a piecewise linear spline, & average rushing yards/attempt with cubic polynomial functional forms. For all association pattern visualizations, please visit the appendices.



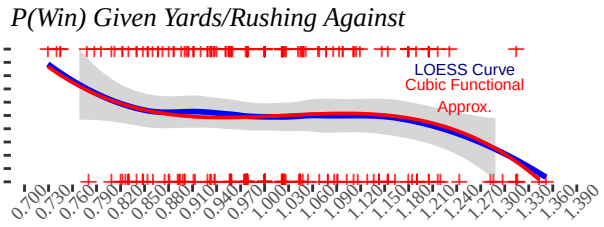








Yards/Rushing Attempt − Team vs Opponent Ratios



Yards/Rushing Against − Team vs Opponent Ratios

Modeling

Aiming to predict NFL game outcomes, it was concluded from correlations that a base model considering Offensive & Defensive Passer Ratings seems both statistically & substantively realistic. The base linear probability model specification reflects both variables' log-functional best approximation. Meanwhile the full model also incorporates **a.** Offensive Sacks together with a dummy indicator, **b.** logged Offensive Rushing Yards/play & piecewise linear spline specifications for **c.** Offensive Turnovers & **d.** Defensive Sacks.

Base vs Full LPMs: Please see these fitted-models' formulaic specifications above. Observing the fitted parameters, one can observe in the Base Model that the intercept is 0.5, implying **a.** if both teams' quarterbacks achieve the same Passer Rating on average throughout the season & **b.** if both teams' opponents' quarterbacks achieve the same passer rating on average against these teams' defenses, than that game is a 50-50 toss-up. Interestingly, Teams whose quarterbacks achieve 1% higher Passer Rating than their opponents can expect to win with 0.83% higher probability, while Teams' defenses against whom quarterbacks achieve 1% higher Passer Rating than their opponents can expect to win with 1.38% lower probability. This suggests it is more important for the team to defend opposing QBs better then the other team, then it is for their QB to outplay the opponents QB, and gives a statistical argument to the football-coach conventional wisdom: "Offense wins Games, Defense wins championships" (though the 2 coefficients are not statistically significantly different). For the coefficients table, please refer to the appendices.

$$P(TeamWin_i = 1)^E = 0.50 + 0.83 * ln(OffQBRating_i) - 1.38 * ln(DefQBRating_i) + \epsilon$$

$$\begin{aligned} P(TeamWin_i = 1)^E = {} & 0.45 + 0.58 * ln(OffQBRating_i) - 0.87 * ln(DefQBRating_i) \\ & -0.09 * ln(OffSacks_i) - 0.41 * D * ln(OffSacks_i) \\ & +0.45 * ln(OffRushYardsAVG_i) - 0.2 * OffTurnovers_i * I(x < 0.8) \\ & +0.11 * OffTurnovers_i * I(0.8 \leq x < 3) - 0.09 * OffTurnovers_i * I(3 \leq x) \\ & +0.15 * DefSacks_i * I(x < 1.6) - 0.14 * DefSacks_i * I(x \leq 1.6) + \epsilon \end{aligned}$$

Stats	LPM Base	LPM Full
adj.r.squared	0.220	0.280
sigma	0.440	0.430
statistic	35.650	9.260
Brier-Score	0.192	0.173
Bias	0.000	0.000
Prediction Acc%	73.800	77.100

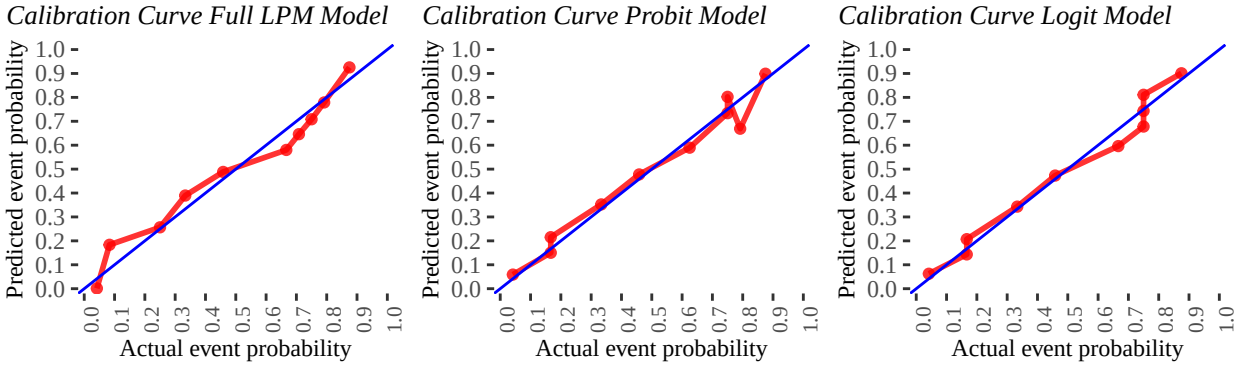
The full model however, seems to outperform the base-model, achieving lower Brier-Scores & 3.3% higher prediction accuracy - obtained via classifying predicted win probabilities with the threshold 0.5. It also

seems more satisfying conceptually, acknowledging that while NFL game outcomes not only depend on the 2 teams' QBs & how well can defenses slow down QBs, it is 2 of the most important determinants. For this reason, I choose the full LPM model to compare it in various forms. According to this model (noting only statistically significant results at 90% confidence):

- Teams whose quarterbacks achieve 10% higher Passer Rating than opponents can expect to win with 5.8% higher probability.
- Teams' defenses against whom quarterbacks achieve 10% higher Passer Rating than opponents can expect to win with 8.7% lower probability
- Teams who achieve 10% higher Average Yards/Rush than opponents can expect to win with 4.5% higher probability.
- Teams' Offenses who suffer 10% more sacks then opponents on average, can expect to win with 0.9% lower probability & if the away team' on average has suffered less sacks than the opponents, but not less then 1/5 of that of the opponent, they can expect to win with 5% lower probability.
- Teams' Offenses who lose the ball 10% more then opponents on average, can expect to win with 1.1% higher probability, assuming the away team' on average lost the ball less than 3 times-, but not less then 4/5th as much as the opponent.
- Teams' Defenses who enforce 10% more sacks then opponents on average, can expect to win with 1.5% higher probability, assuming the away team' on average enforced not more then 1.6 times as many sacks as their opponent.

Full LPM vs Logit & Probit Models: Comparing Model summary statistics to using the Logistic & Gaussian Cumulative Density Functions as link-functions improves the Brier-score minimally, & achieves lower Akaike & Bayesian Information Criterion scores, even though resulting in 1.3% loss of accuracy & very slight Bias for the Probit model. Comparing LPM coefficents & the Probit & Logit models' marginal effects, minimal differences are observable, largest being -0.07, for ln(DefQBRating), which implies the difference in expected win probability associated with a 10% difference between 2 teams' defensive performance vs quarterbacks is 0.7% model-to-model, indeed minimal. Though the Logit model boasts lower Brier-score & less bias, the Probit Model almost perfectly calibrated between the 20th & 80th percentile of predictions, & is therefore chosen as the final model to predict NFL game outcomes with.

Stats	LPMFull	Probit	Logit
logLik	-129.880	-123.820	-123.360
AIC	285.760	271.650	270.720
BIC	331.010	313.410	312.490
Bias	0.000	-0.001	0.000
Brier-Score	0.173	0.169	0.168
Prediction Acc%	77.100	75.800	75.800



Robustness Check: As part of the effort to generalize association patterns between various NFL game statistics & game outcomes, the obvious counter-argument with the model above is the fact that it uses data from 2020. Covid-19 continues to impact every aspect of life & the NFL is no different. Games are played behind closed doors, teams could not hold similar training camps vs previous years (e.g. there were no pre-season games) & any players can become unavailable to play due to getting infected. The former may be a reason why no bias towards home team wins is observable, while lack of training camp & the risk of getting infected may lead to less predictability. These factors warrant to re-estimate the chosen probit model using data from the 2019 Season, to compare deviations in coefficients. Based on the table below, apart from the 2 Passer Rating related variables, no explanatory variable remained statistically significant at 10% significance, suggesting the model-formula as specified above not to be robust. the 2 passer rating variables however are robust, as they do not differ statistically significantly 10% significance. On another note the model remained unbiased, though boasting 6.2% lower prediction accuracy.

	MgEff	SE	P_Val	MgEffly	SEly	P_ly	Summ Stats	Probit LY	Probit
ln(OffQBRating)	0.54	0.16	0.00	0.77	0.24	0.00	Null.Dev	332.11	332.69
ln(DefQBRating)	-0.80	0.33	0.02	-1.03	0.23	0.00	df.null	239.00	239.00
OffSackDummy	0.07	0.06	0.29	-0.02	0.14	0.90	logLik	-129.30	-123.82
ln(OffSacks)	-0.07	0.12	0.55	-0.06	0.27	0.83	AIC	282.60	271.65
ln(OffRushYardsAVG)	0.42	0.22	0.06	0.15	0.20	0.46	BIC	324.37	313.41
PLS(OffTOs)-<0.8	-0.20	0.17	0.24	0.25	0.17	0.15	deviance	258.60	247.65
PLS(OffTOs)-0.8< x <3	0.11	0.06	0.07	-0.01	0.05	0.91	df.res	228.00	228.00
PLS(OffTOs)>3	-0.08	0.06	0.15	-0.09	0.06	0.12	NrObs	240.00	240.00
PLS(DefSacks)<1.6	0.13	0.08	0.11	0.02	0.13	0.88	Bias	0.00	0.00
PLS(DefSacks)>1.6	-0.13	0.14	0.37	0.31	0.66	0.64	Pred Acc%	69.60	75.80
OffSackDummy*ln(OffSacks)	-0.43	0.15	0.00	0.02	0.29	0.95	Brier Sc	0.18	0.17

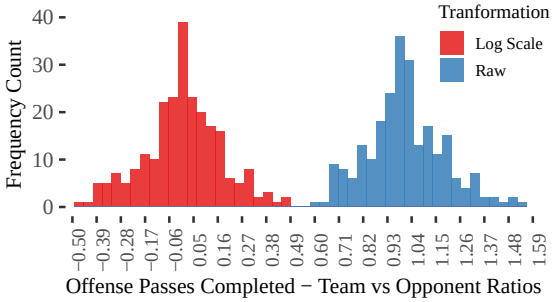
External Validity: The final consideration is to test model performance on a test sample, for comparability with previous research. To do so, Predictive Accuracy, Bias & the Brier Score are computed on the aforementioned data from the 2019 season. Note, for Predictive Accuracy a threshold value of 0.5 is used to obtain game outcomes for win probabilities & no distinction is made between false-positives & negatives, as there is no material difference for it for gambling purposes. As can be seen below, the model remained unbiased, yet achieved 6.25% lower accuracy during testing then on training data, incidentally exactly 1 less game per week. The 69.58% accuracy equals ca. 11/16 games (on an average week). This seems to rival recent NFL game prediction efforts, with Warner (2010) achieving 64.4% with a Gaussian process model using 10 years' data, Shau (2011) achieving 68.4% with Support Vector Machine (SVM) model using 41 years' data, & Owen & Galle (2014) achieving 68.6%, also with SVM & 5 years' data.

Season	Sample	P(Win)	Pred(Win)	PredAcc%	Brier	Bias
2020	Test	0.475	0.477	69.58	0.181	0.002
2019	Training	0.496	0.495	75.83	0.169	-0.001

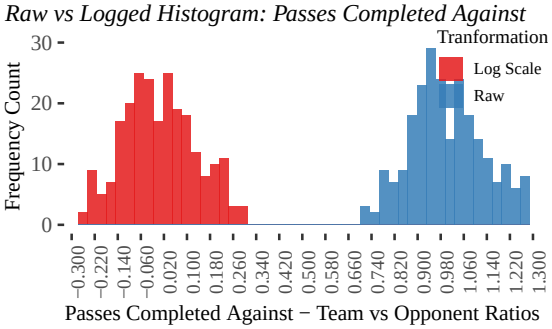
Appendices

1) Raw & Log-Transformed Histograms

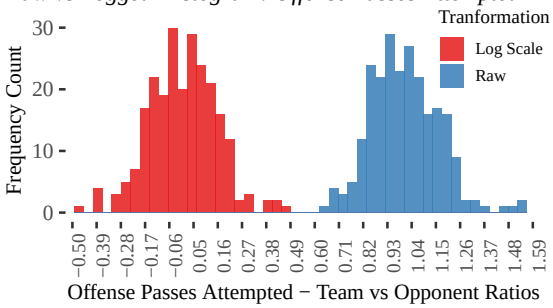
Raw vs Logged Histogram: Offense Passes Completed



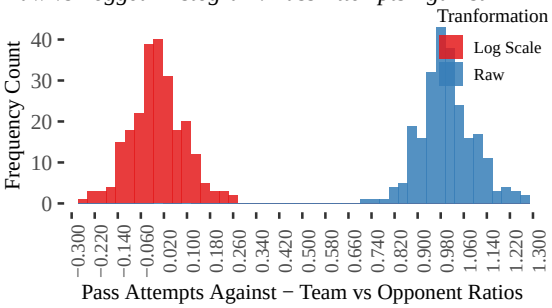
Raw vs Logged Histogram: Passes Completed Against



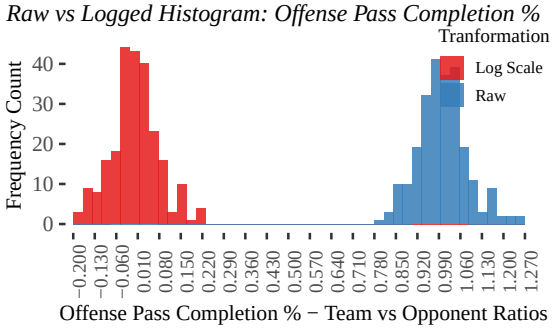
Raw vs Logged Histogram: Offense Passes Attempted



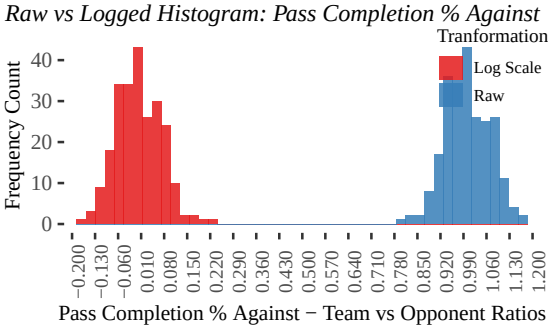
Raw vs Logged Histogram: Pass Attempts Against



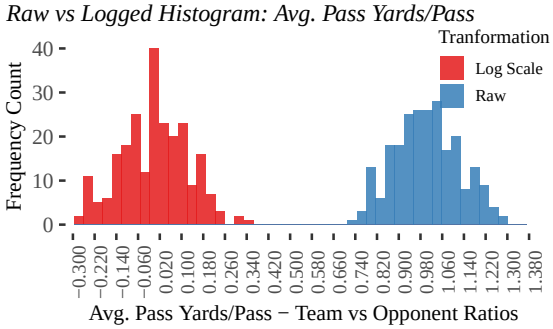
Raw vs Logged Histogram: Offense Pass Completion %



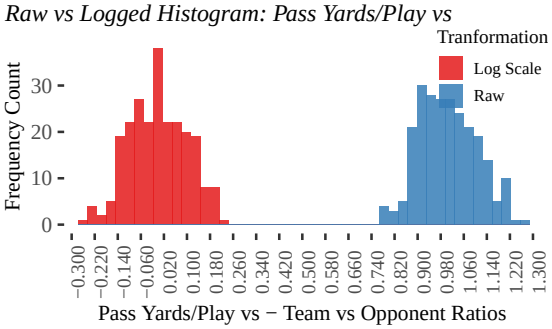
Raw vs Logged Histogram: Pass Completion % Against



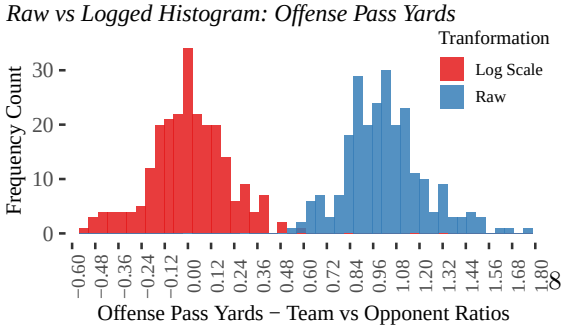
Raw vs Logged Histogram: Avg. Pass Yards/Pass



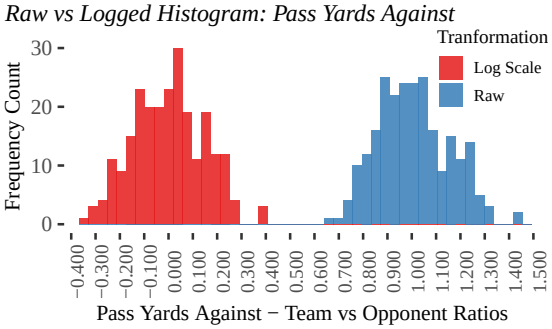
Raw vs Logged Histogram: Pass Yards/Play vs



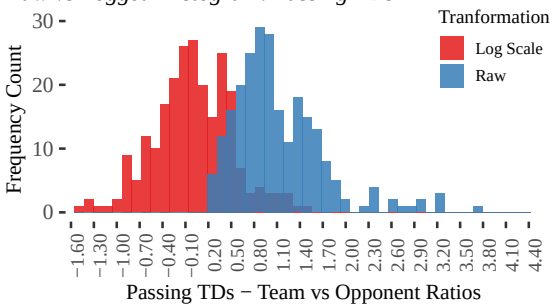
Raw vs Logged Histogram: Offense Pass Yards



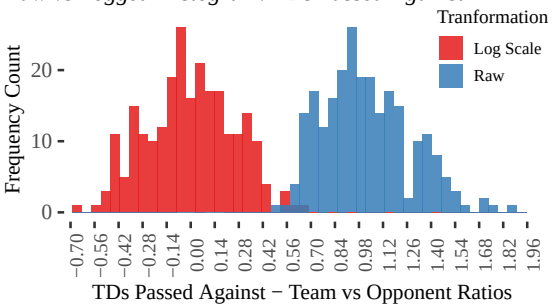
Raw vs Logged Histogram: Pass Yards Against



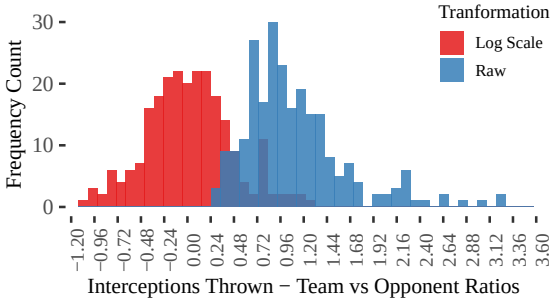
Raw vs Logged Histogram: Passing TDs



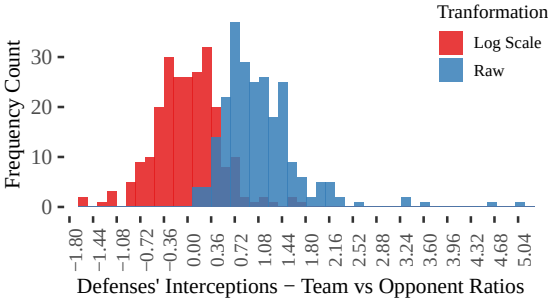
Raw vs Logged Histogram: TDs Passed Against



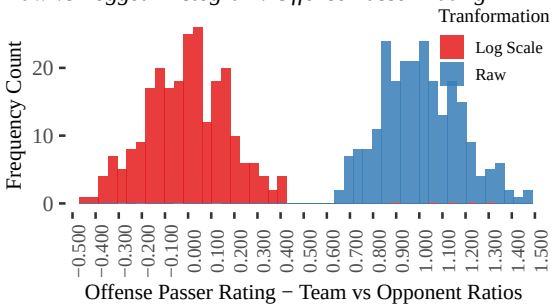
Raw vs Logged Histogram: Interceptions Thrown



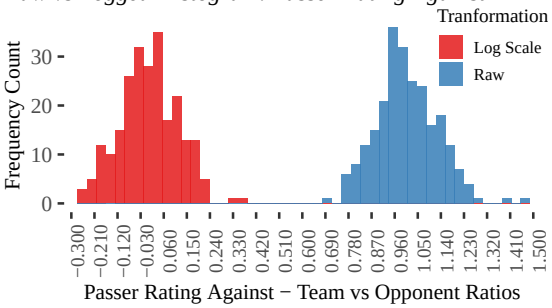
Raw vs Logged Histogram: Defenses' Interceptions



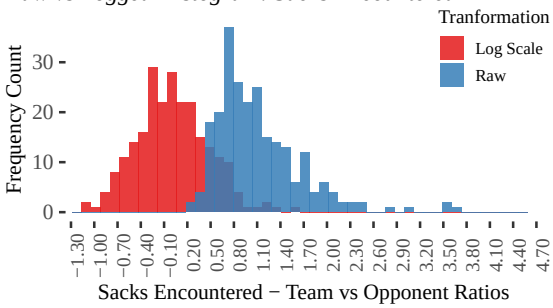
Raw vs Logged Histogram: Offense Passer Rating



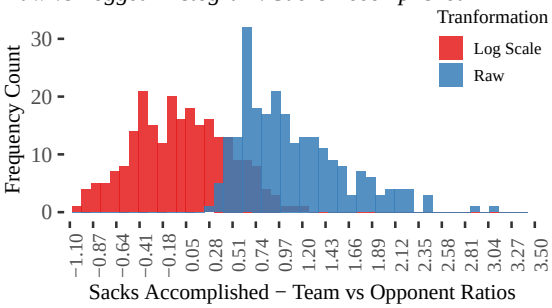
Raw vs Logged Histogram: Passer Rating Against



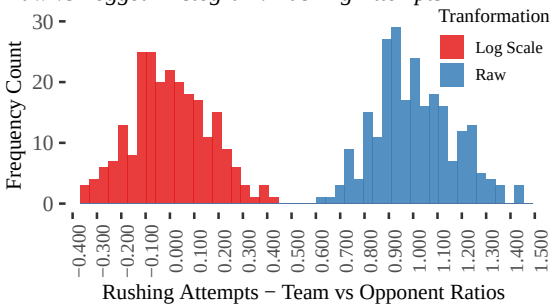
Raw vs Logged Histogram: Sacks Encountered



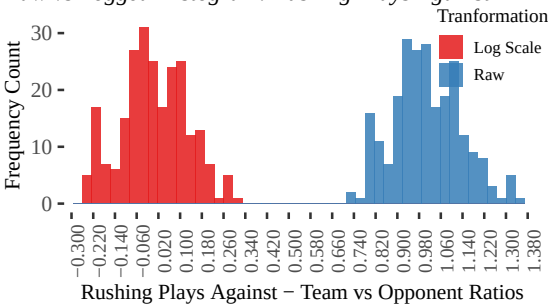
Raw vs Logged Histogram: Sacks Accomplished



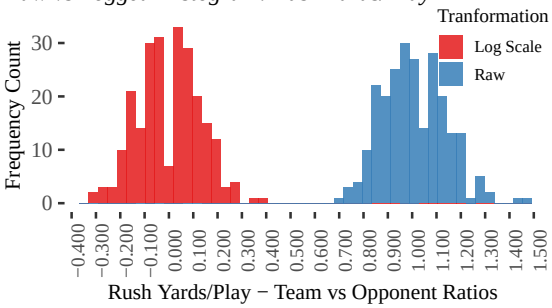
Raw vs Logged Histogram: Rushing Attempts



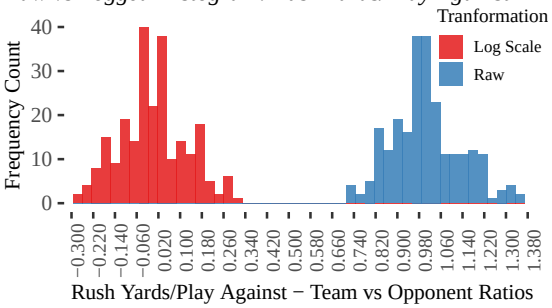
Raw vs Logged Histogram: Rushing Plays Against



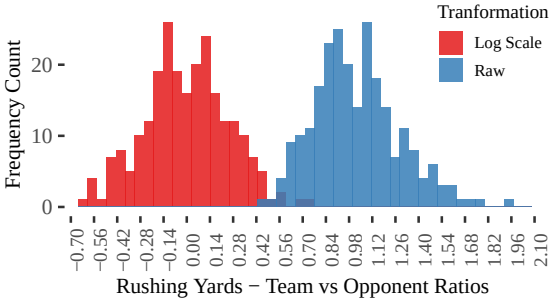
Raw vs Logged Histogram: Rush Yards/Play



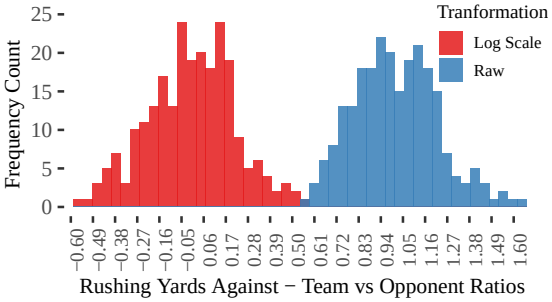
Raw vs Logged Histogram: Rush Yards/Play Against



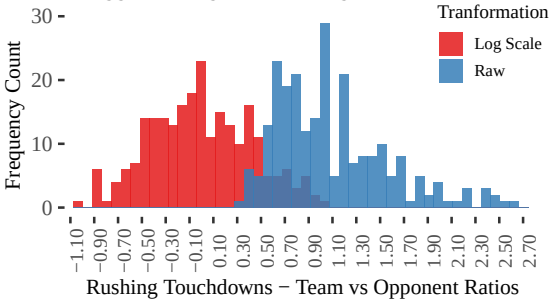
Raw vs Logged Histogram: Rushing Yards



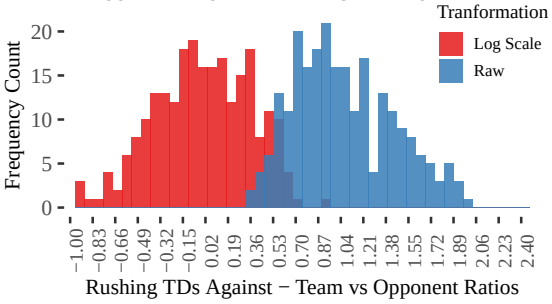
Raw vs Logged Histogram: Rushing Yards Against



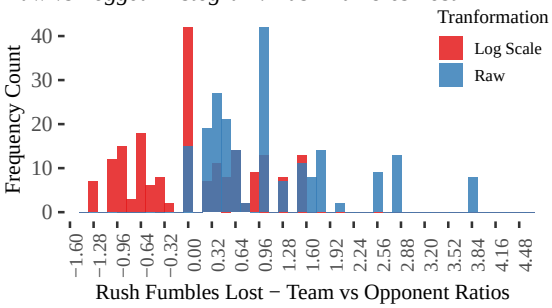
Raw vs Logged Histogram: Rushing Touchdowns



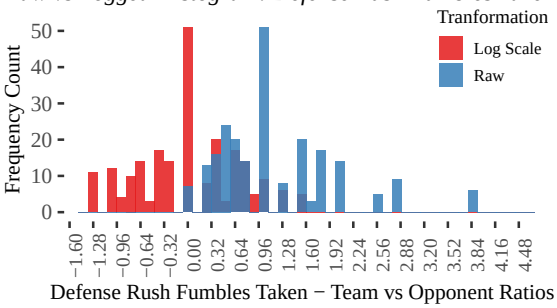
Raw vs Logged Histogram: Rushing TDs Against



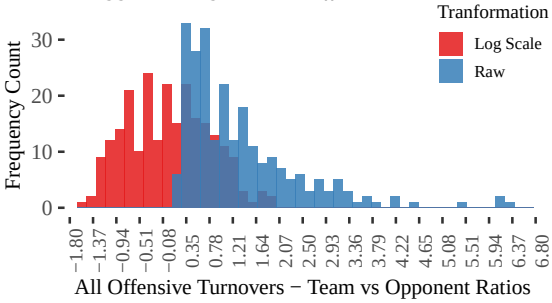
Raw vs Logged Histogram: Rush Fumbles Lost



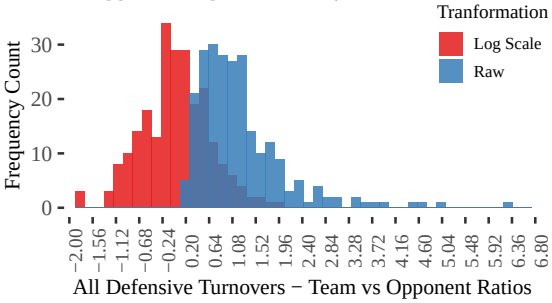
Raw vs Logged Histogram: Defense Rush Fumbles Taken



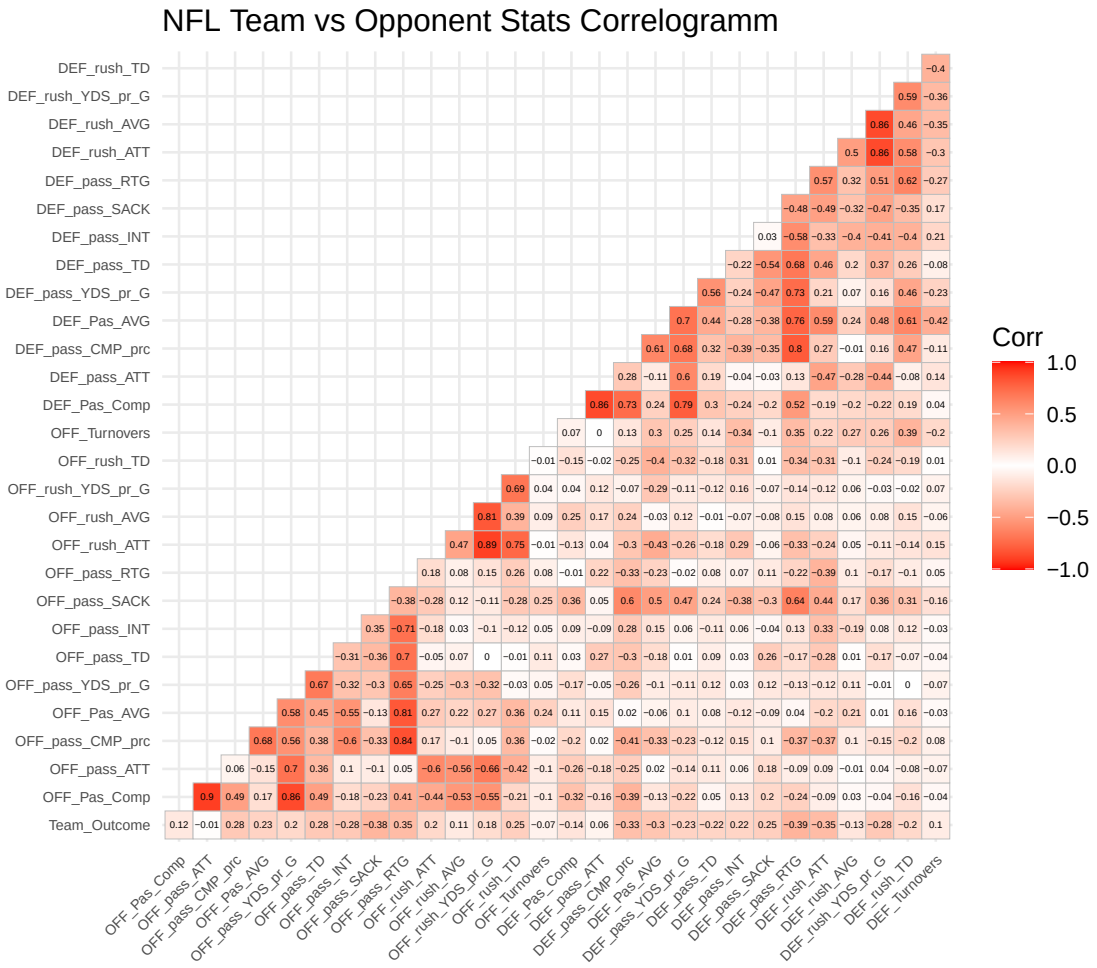
Raw vs Logged Histogram: All Offensive Turnovers



Raw vs Logged Histogram: All Defensive Turnovers

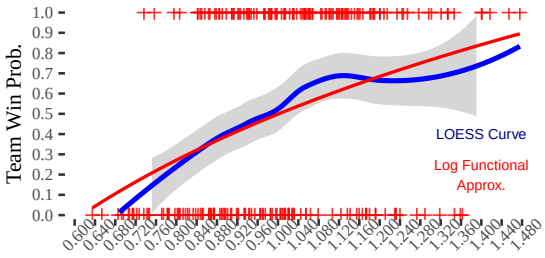


2) Variable Correlogram



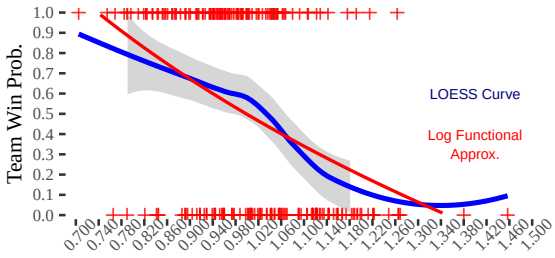
3) Association Patterns & Functional Approximations

P(Win) Given Passer Rating



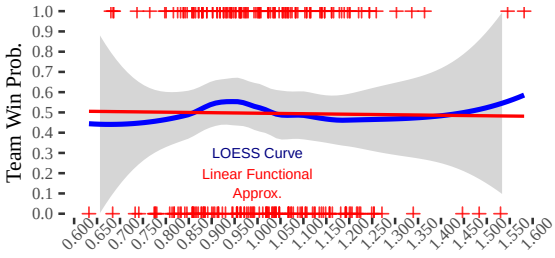
Passer Rating − Team vs Opponent Ratios

P(Win) Given Opposing Passers' Rating



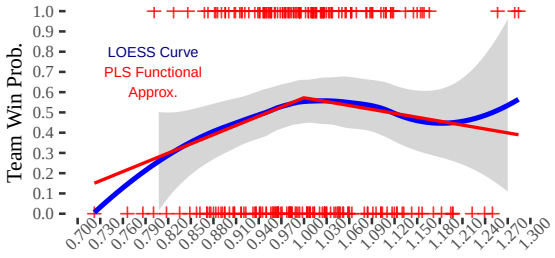
Opposing Passers' Rating − Team vs Opponent Ratios

P(Win) Given Offensive Pass Attempts



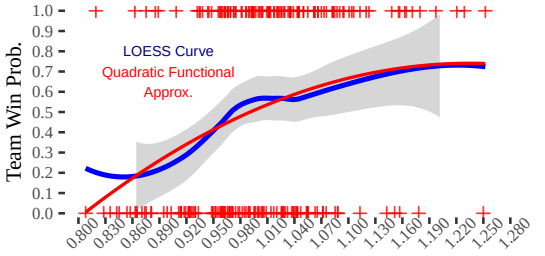
Offensive Pass Attempts − Team vs Opponent Ratios

P(Win) Given Passes Attempted Against



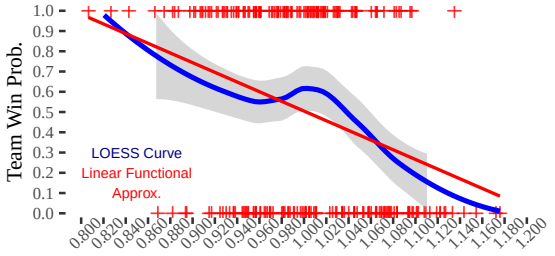
Passes Attempted Against − Team vs Opponent Ratios

P(Win) Given Pass Completion %



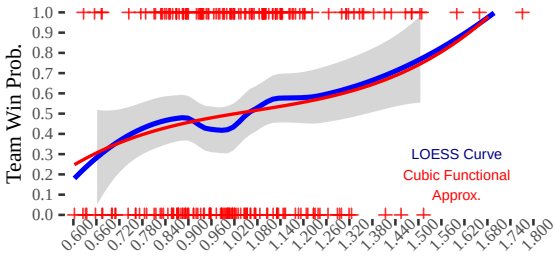
Pass Completion % − Team vs Opponent Ratios

P(Win) Given Pass Completion % Against



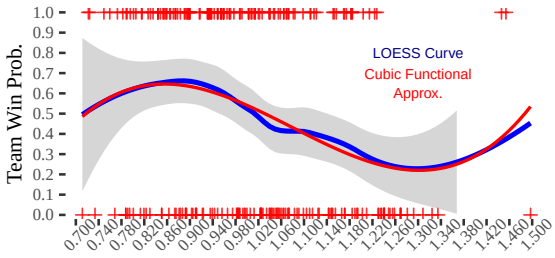
Pass Completion % Against − Team vs Opponent Ratios

P(Win) Given Passing Yards/Game



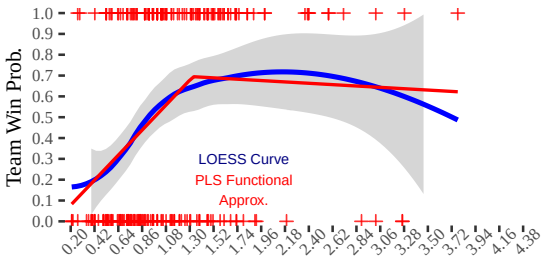
Passing Yards/Game − Team vs Opponent Ratios

P(Win) Given Passing Yards/Game Against



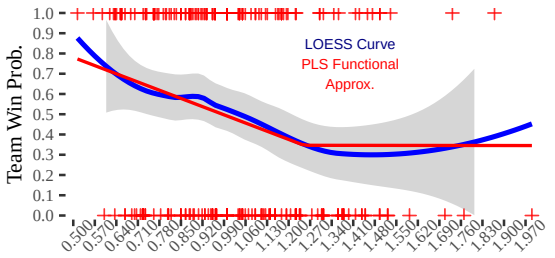
Passing Yards/Game Against − Team vs Opponent Ratios

P(Win) Given PLS of Passing TDs

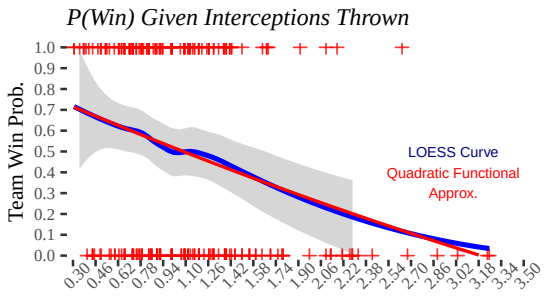


PLS of Passing TDs − Team vs Opponent Ratios

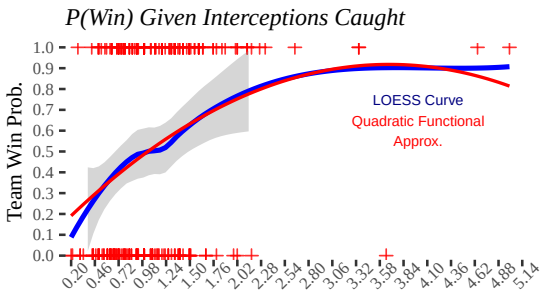
P(Win) Given PLS of TD Passes Against



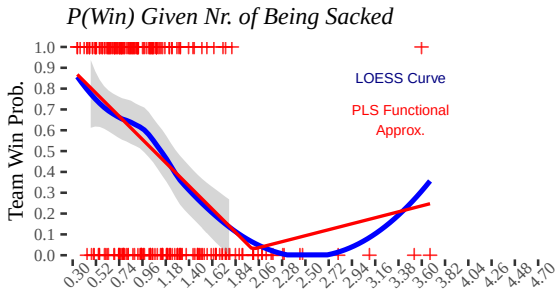
PLS of TD Passes Against − Team vs Opponent Ratios



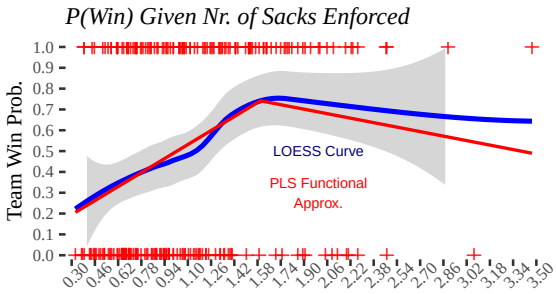
Interceptions Thrown − Team vs Opponent Ratios



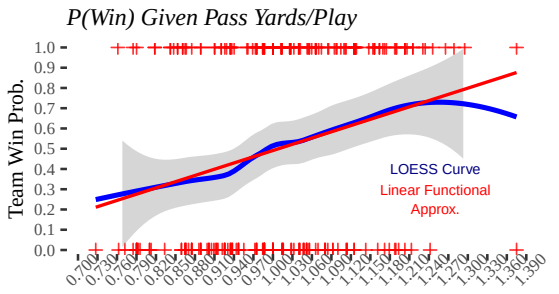
Interceptions Caught − Team vs Opponent Ratios



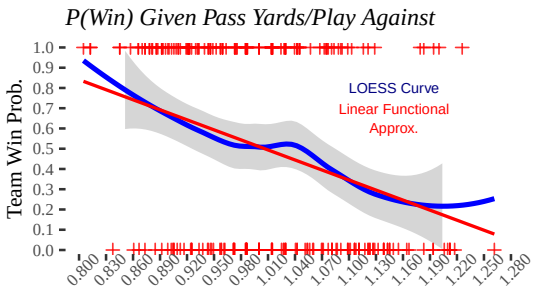
Nr. of Being Sacked − Team vs Opponent Ratios



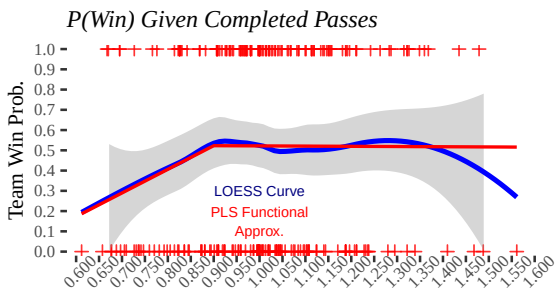
Nr. of Sacks Enforced − Team vs Opponent Ratios



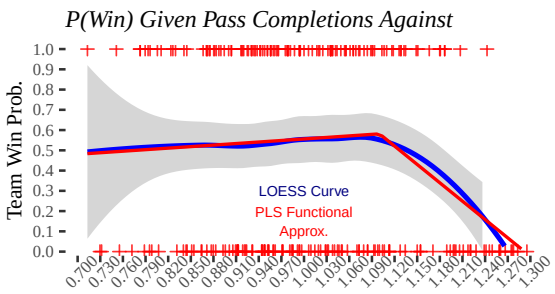
Pass Yards/Play − Team vs Opponent Ratios



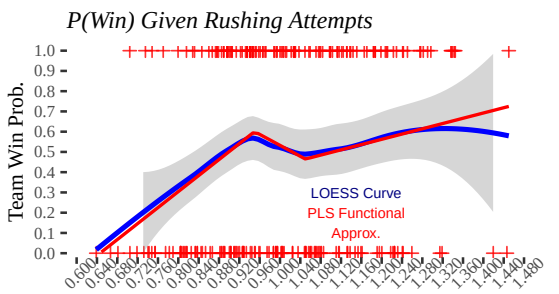
Pass Yards/Play Against − Team vs Opponent Ratios



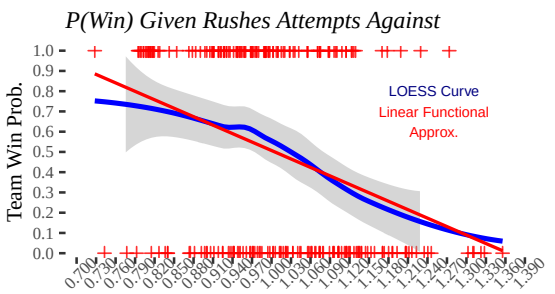
Completed Passes − Team vs Opponent Ratios



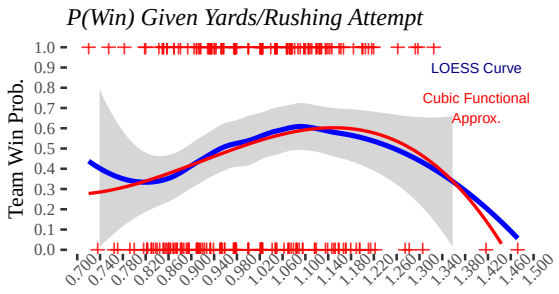
Pass Completions Against − Team vs Opponent Ratios



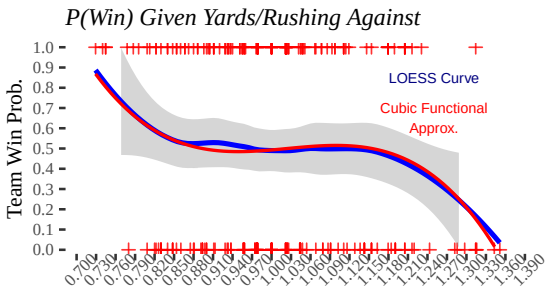
Rushing Attempts − Team vs Opponent Ratios



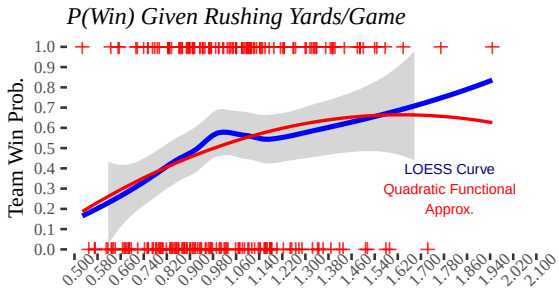
Rushes Attempts Against − Team vs Opponent Ratios



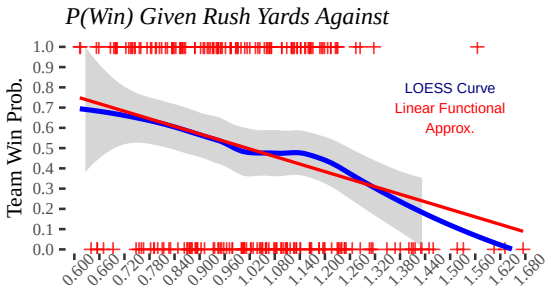
Yards/Rushing Attempt − Team vs Opponent Ratios



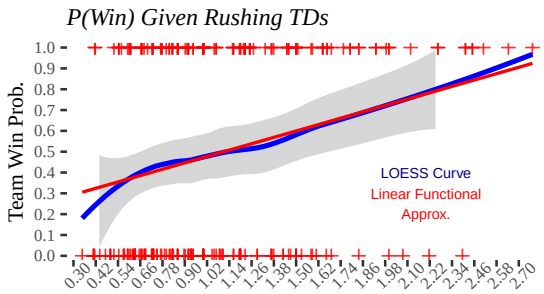
Yards/Rushing Against − Team vs Opponent Ratios



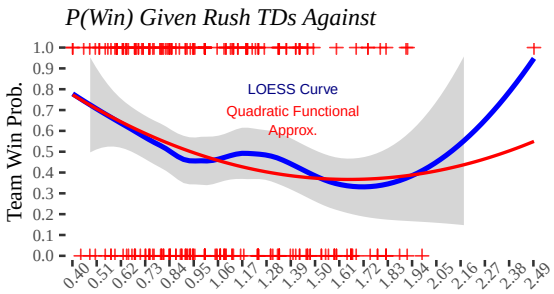
Rushing Yards/Game − Team vs Opponent Ratios



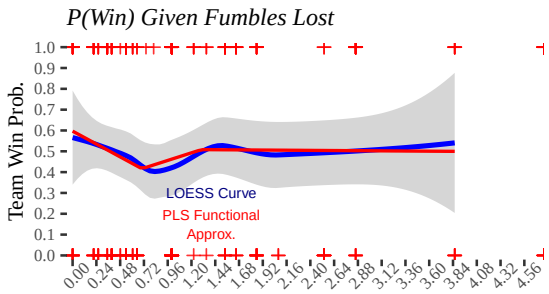
Rush Yards Against − Team vs Opponent Ratios



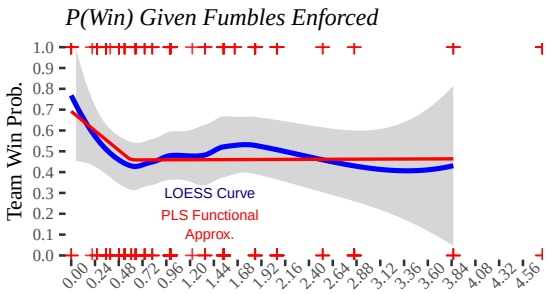
Rushing TDs − Team vs Opponent Ratios



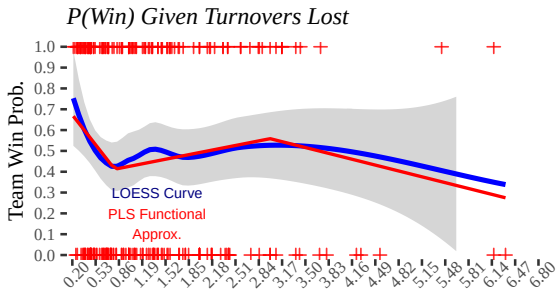
Rush TDs Against − Team vs Opponent Ratios



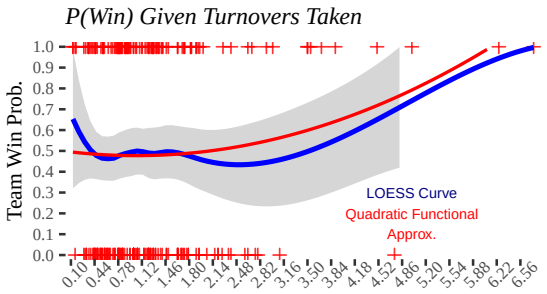
Fumbles Lost − Team vs Opponent Ratios



Fumbles Enforced − Team vs Opponent Ratios



Turnovers Lost − Team vs Opponent Ratios

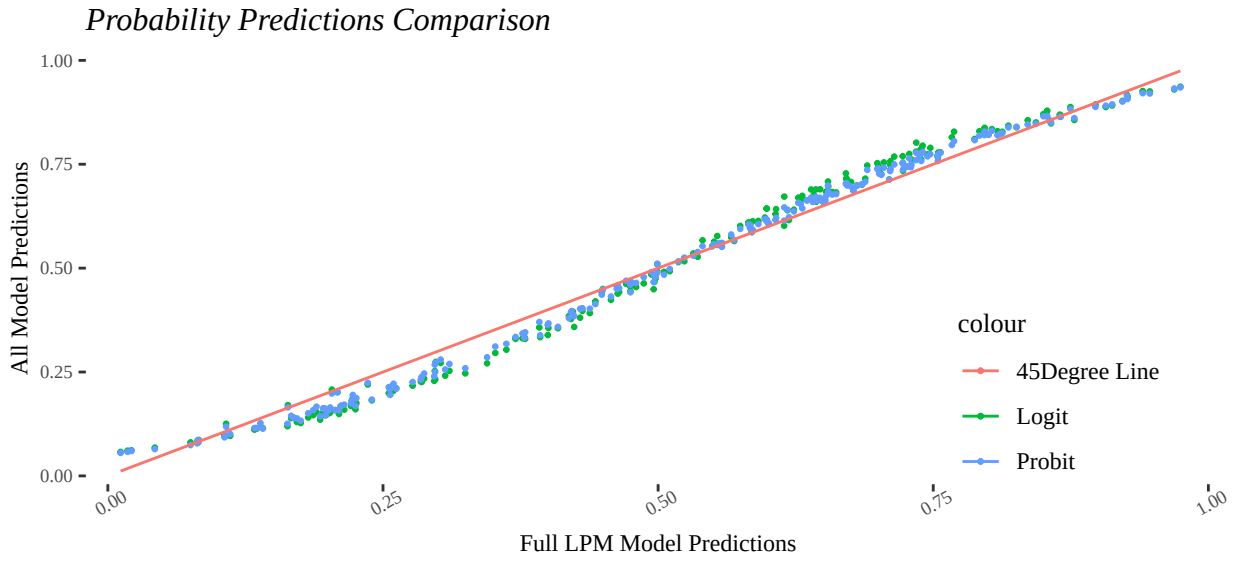


Turnovers Taken − Team vs Opponent Ratios

4) Base vs Full LPM Model Coefficients table

	Estimate	Std. Error	t value	Pr(>|t|)	Model
Constant	0.5004	0.0285	17.5681	0	Base
ln(OffQBRating)	0.8255	0.1656	4.9853	0	Base
ln(DefQBRating)	-1.3741	0.2491	-5.517	0	Base
Constant1	0.4457	0.1565	2.8474	0.0048	Full
ln(OffQBRating)1	0.5764	0.1838	3.1361	0.0019	Full
ln(DefQBRating)1	-0.8708	0.3516	-2.477	0.014	Full
ln(OffRushRardsAVG)	0.4481	0.2276	1.9688	0.0502	Full
PLS(OffTurnovers)-<0.8	-0.2037	0.1839	-1.1079	0.2691	Full
PLS(OffTurnovers)-0.8< x <3	0.1107	0.0517	2.1386	0.0335	Full
PLS(OffTurnovers)>3	-0.0922	0.066	-1.3982	0.1634	Full
PLS(DefSacks)<1.6	0.1517	0.0889	1.7065	0.0893	Full
PLS(DefSacks)>1.6	-0.1399	0.1389	-1.0075	0.3148	Full
OffSackDummy * ln(OffSacks)	-0.4074	0.1547	-2.6339	0.009	Full

5) LPM Coefficients, Logit & Probit Marginal Effects table

Variables	LPM_Coeffs	LPM_SE	Probit_M_Effs	Probit_SE	Logit_M_Effs	Logit_SE
Constant	0.45	0.16	NA	NA	NA	NA
ln(OffQBRating)	0.58	0.18	0.54	0.16	0.55	0.20
ln(DefQBRating)	-0.87	0.35	-0.80	0.33	-0.82	0.37
OffSackDummy	0.07	0.07	0.07	0.06	0.07	0.06
ln(OffSacks)	-0.09	0.11	-0.07	0.12	-0.06	0.12
ln(OffRushYardsAVG)	0.45	0.23	0.42	0.22	0.43	0.24
PLS(OffTurnovers)-<0.8	-0.20	0.18	-0.20	0.17	-0.20	0.18
PLS(OffTurnovers)-0.8< x <3	0.11	0.05	0.11	0.06	0.11	0.06
PLS(OffTurnovers)>3	-0.09	0.07	-0.08	0.06	-0.09	0.06
PLS(DefSacks)<1.6	0.15	0.09	0.13	0.08	0.13	0.09
PLS(DefSacks)>1.6	-0.14	0.14	-0.13	0.14	-0.13	0.14
OffSackDummy * ln(OffSacks)	-0.41	0.15	-0.43	0.15	-0.44	0.18

6) Predicted Probabilities Comparison - LPM Full, Logit & Probit models