Journal of Quantitative Analysis in Sports

Manuscript 1210

Rush versus Pass: Modeling the NFL

Ralph Abbey[††] John Holodnak[*] Chandler May[†]
Carl Meyer[††] Dan Moeller[‡]

†† North Carolina State University

* Ohio Northern University

† Harvey Mudd College

‡ University of Notre Dame

1 Introduction

A common question in football is whether a strong rushing or passing offense is more important in determining the outcome of a game. On the surface, it is easy to see both sides of the debate. A powerful running game tends to slowly and deliberately advance the ball down the field, using large amounts of time, while a strong passing game can produce large gains and high scores. We look at the question from the perspective of predicting the outcome of National Football League (NFL) games based on the relative strengths of teams' rushing or passing attacks. These relative strengths are derived from four sports ranking systems: the Keener (1993) ranking model, the Massey (1997) Least Squares model, the Govan et al. (2009) Offense-Defense model, and the Generalized Markov model (Govan, 2008).

We also consider the correlation between differences in rushing or passing yards and differences in score. After all, if either statistic is a good predictor of game outcomes, then outgaining an opponent in that statistic should correspond to outscoring an opponent. We assume that a statistic that correlates well with scores is a good indicator of team strength.

Finally, we examine the fractions of games won by teams that outgained their opponents in rushing yards, passing yards, or combinations of the two. These fractions are akin to the observational analogues of conditional probabilities.

2 Background

Two common terms associated with sports models are ranking and rating; each will be used often throughout this paper. A ranking of N teams places them in order of relative importance, with the best team receiving rank one. A rating of the same teams describes the degree of relative importance of each team.

It will also be useful to define rushing and passing yards. Total rushing yards is simply the sum of the yards gained on each rushing play, remembering that the number of yards gained on a particular play may be negative. Total passing yards is the sum of the yards gained on each forward passing play minus the number of yards lost through sacks of the quarterback. Again, a passing play may result in negative yardage. We make the distinction of forward passing play to distinguish from a pitch or pass in which the receiver is further back on the field than the quarterback, both of which count towards rushing yardage.

Many of our results are based on "foresight prediction," that is, prediction of the outcome of the games in each week, using the data from the previous weeks. For example, to predict the outcome of games in week five, we load the data from the first four weeks of the season and use the rankings produced by the models to predict game outcomes. The "foresight accuracy" for a week is the percentage of games predicted correctly in that week. The foresight accuracy for a season is the weighted average (weeks have different numbers of games) of each of the foresight accuracies from the individual weeks.

3 Summary of the Models

3.1 Keener's Ranking Method

The Keener (1993) ranking model makes two fundamental assumptions. His first assumption is that the *strength* of a team is based upon its interactions with opponents. He defines the strength of team i to be

$$s_i = \frac{1}{n_i} \sum_{j=1}^{N} a_{ij} r_j \tag{1}$$

where a_{ij} is a non-negative value that depends upon the outcome of the game between i and j, r_j is the rating of team j, n_i is the number of games played by team i, and N is the total number of teams. Keener assigns the value of a_{ij} as follows

$$a_{ij} = h(\frac{S_{ij} + 1}{S_{ij} + S_{ji} + 2}) \tag{2}$$

$$h(x) = \frac{1}{2} + \frac{1}{2}\text{sgn}(x - \frac{1}{2})\sqrt{|2x - 1|} \tag{3}$$

where S_{ij} is the number of points i scored against j. To clarify, if teams i and j play more than one game, S_{ij} represents the total number of points scored by i against j. The assignment of a_{ij} is complicated, but the underlying logic is not. Essentially, we want to split the reward for competing between the two teams. In close games, the split will be fairly even. (Note that in a tie, the split is exactly half and half). The purpose of the non-linear h function is to minimize the incentive of the winning team to "run up" the score. The h function also has the effect of widening the reward gap in close games.

Keener's second assumption is that the *strength* of a team should be proportional to a team's rating. In equation form

$$\mathbf{s} = \mathbf{A}\mathbf{r} = \lambda\mathbf{r} \tag{4}$$

where $\mathbf{A}$ is an $N \times N$ matrix with a_{ij} as components, $\mathbf{r}$ is a column vector of N ratings, and $\mathbf{s}$ is a column vector of N strengths. By the Perron-Frobenius Theorem (Meyer, 2000), this eigenvalue-eigenvector equation will have a unique, up to a scalar multiple, and positive solution, provided that the matrix $\mathbf{A}$ is non-negative and irreducible. Due to the limited number of games played, it is likely that the matrix will be reducible during the early weeks of the season. To ensure irreducibility, we slightly perturb the matrix

$$\mathbf{A}_p = \mathbf{A} + \varepsilon \mathbf{e}\mathbf{e}^T. \tag{5}$$

After calculating $\mathbf{r}$, we use the ratings to create a ranking of teams that can be used to predict game outcomes.

3.2 Massey Least Squares Model

The Massey (1997) model makes one fundamental assumption. Namely, the difference in team's ratings should be proportional to the difference in points scored

$$r_i - r_j = y_k \tag{6}$$

where r_i is the rating of the i^{th} team and y is the difference in points scored in the game between team i and team j. For simplicity, the constant of proportionality is assumed to be one. A system of such equations easily admits itself to matrix form as follows

$$\mathbf{X}\mathbf{r} = \mathbf{y}, \tag{7}$$

where $\mathbf{X}$ is a $K \times N$ matrix (K is the total number of games), and where the k^{th} row contains a one in the column corresponding to the winning team and a negative one in the column corresponding to the losing team. Additionally, $\mathbf{r}$ is a ratings vector with r_i the rating of the i^{th} team and $\mathbf{y}$ a vector of point differences where y_k is the point differential in the k^{th} game. To be clear, $\mathbf{r}$ is a column vector of N ratings and $\mathbf{y}$ is a column vector of K point differentials.

In most practical applications, the system of equations will be overdetermined. For example in the NFL, there are 267 games in the season and only 32 teams; thus, the system will become overdetermined after only a few weeks of data is loaded. The strategy is to look for the least squares solution to the system

$$\mathbf{X}^T\mathbf{X}\mathbf{r} = \mathbf{X}^T\mathbf{y}. \tag{8}$$

The question we must now answer is whether or not Massey's rating vector $\mathbf{r}$ is unique. Unfortunately, due to the fact that each row sum is zero, the columns of $\mathbf{X}$ are not linearly independent and thus $\mathbf{X}$ does not have full rank. Therefore, there is no unique solution to the least squares problem. Provided, however, that the matrix is saturated ($\mathbf{e} = \begin{bmatrix}1 & 1 & \cdots & 1\end{bmatrix}^T$ is the only nontrivial vector in the nullspace), full rank can be obtained by adding an additional condition to the system. The simplest solution is to require the rating vector to sum to zero by appending a row of ones to $\mathbf{X}$ and a zero to $\mathbf{y}$.

3.3 Offense-Defense Model

The Offense-Defense model (Govan, 2008) defines offensive rating o_i and defensive rating d_i of team i as follows

$$o_j = m_{1j}\frac{1}{d_1} + m_{2j}\frac{1}{d_2} + \cdots + m_{nj}\frac{1}{d_n} \tag{9}$$

$$d_i = m_{i1}\frac{1}{o_1} + m_{i2}\frac{1}{o_2} + \cdots + m_{in}\frac{1}{o_n} \tag{10}$$

where m_{ij} is the number of points scored by j against i. The offensive formula sums the points scored by a particular team's offense on each team played, dividing by the defensive ratings of the corresponding teams. Similarly, the defensive rating sums the points allowed by a particular defense, dividing by the offensive ratings of the corresponding teams. A large offensive and small defensive rating are considered "good." Thus, as the number of points scored on a team is divided by the defensive rating of that team, scoring a large number of points against a good defense will do more for a team's offensive rating than scoring the same number of points on a weak defense. Likewise, holding a good offense to a few points is better for a team's defensive rating than holding a poor offense to a few points. We can aggregate the two ratings with the ratio $\frac{o_i}{d_i}$, which maintains our intuitive belief that a bigger rating is better.

If we now let $\mathbf{o}$ be a column vector of offensive ratings, $\mathbf{d}$ be a column vector of defensive ratings, and $\mathbf{M}$ be the $N \times N$ matrix with m_{ij} as entries, then these formulae can be expressed recursively for $k = 1, 2, \ldots$ as

$$\mathbf{o}^{(k)} = \mathbf{M}^T \frac{1}{\mathbf{d}^{(k-1)}} \tag{11}$$

$$\mathbf{d}^{(k)} = \mathbf{M}\frac{1}{\mathbf{o}^{(k)}} \tag{12}$$

where $\frac{1}{\mathbf{d}}$ and $\frac{1}{\mathbf{o}}$ are the elementwise inverses of $\mathbf{d}$ and $\mathbf{o}$, and where we initialize $\mathbf{d}^{(0)} = \begin{bmatrix}1 & 1 & \cdots & 1\end{bmatrix}^T$.

These recursive formulae are equivalent to a row-column stochastic balancing of **M**. Thus, as a result of the Sinkhorn-Knopp Theorem (Sinkhorn and Knopp, 1967), we know that these formulae will converge as k approaches infinity. The Sinkhorn-Knopp Theorem requires, however, that the matrix **M** have total support. A nonnegative $N \times N$ matrix **B** with elements b_{ij} has *total support* if, for every positive element b_{ij}, there is some permutation σ of the numbers $1, \ldots, N$ such that $\sigma(i) = j$ and every element of the set $\{b_{1\sigma(1)}, \ldots, b_{N\sigma(N)}\}$ is positive. To ensure total support, we slightly perturb the matrix

$$\mathbf{M}_p = \mathbf{M} + \varepsilon \mathbf{e}\mathbf{e}^T. \tag{13}$$

3.4 Generalized Markov Model

The Generalized Markov model (Govan, 2008) constructs an $N \times N$ matrix **S** where s_{ij} is the sum of the score differences in each game that i lost to j. We then normalize **S** to make it row stochastic. We can imagine a directed graph, with teams as nodes and directed edges that are the positive normalized point spreads. The edges point from loser to winner. The weight of an edge will be the total loss margin from all games that i lost to j, divided by the total loss margin for all games that i lost. The interpretation of this graph and corresponding matrix is that a team "votes for" the teams to which it lost. Thus, a team with many losses will vote for many other teams and vote most towards the teams that beat it by the largest margin, while a team with many wins will have many teams voting for it.

The real strength of the Generalized Markov model is that it can input several statistics at once, building several **S** matrices and summing them in a convex combination, as in

$$\mathbf{G} = \alpha_1 \mathbf{S_1} + \alpha_2 \mathbf{S_2} + \cdots + \alpha_n \mathbf{S_n} \tag{14}$$

where, for example, $\mathbf{S}_1$ could be the stochastic matrix of score differences, $\mathbf{S}_2$ could be the stochastic matrix of passing yard differences, and so on. Since **G** is the convex combination of stochastic matrices, it will also be stochastic. Hence, it will have a unique, up to a scalar multiple, and positive left eigenvector. This left eigenvector is the limiting probability vector and also our ratings vector.

3.5 Model Validity

We assert the validity of these models, with scores as the input, by comparing their game prediction accuracies in the 2008 NFL season with the accuracies of two ESPN analysts, Chris Mortensen and Mike Golic (ESPN, 2009). This comparison is displayed in Table 1.

Keener	Massey	Offense-Defense	Gen. Markov	Mortensen	Golic
0.628	0.637	0.630	0.597	0.650	0.609

Table 1: Foresight Accuracies of Models and ESPN Analysts

4 Rushing and Passing Yards as Indicators of Team Strength

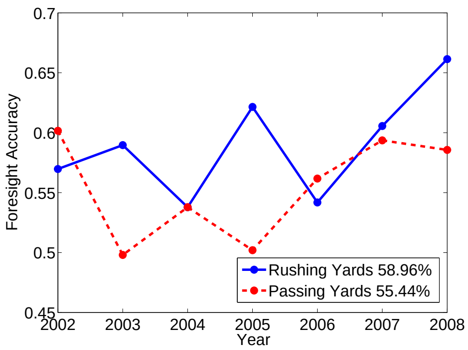
4.1 Foresight Accuracy in the Models

All of the models discussed above use scores as their primary input. Since our purpose is to compare the relative effectiveness of rushing and passing yards at predicting game outcomes, we will load these statistics into the models instead of scores. To be clear, our goal is to determine whether rushing or passing is a better game predictor and thus, we believe, a better indicator of team strength. We do not claim that rushing or passing yards is a superior indicator in comparison with scores; we seek only to compare the two.

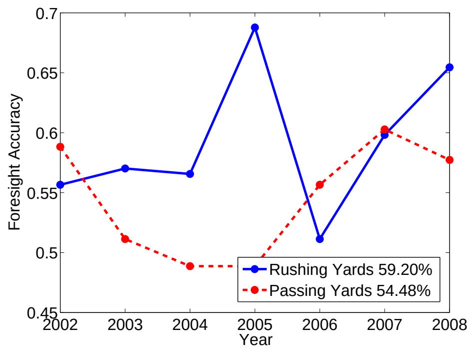
Changing the statistic used in the models does not distort their original intentions. For example, if we change the Offense-Defense model to accept rushing yards, the offensive rating of a team becomes the sum, over all possible opponents, of the rushing yards the team gained against an opponent divided by the opponent's defensive rating. This is still a logical way of measuring the team's offensive strength. The Generalized Markov model is obviously configured to accept any sort of statistic and the Keener model readily adapts as well. There is only the rare case when a statistic turns out to be negative. We identified four cases in our rushing and passing yard data, for example, when a team actually concluded a game with negative total rushing or passing yards. Since non-negativity of the matrix is required for several of our theorems to apply, we merely set these negative values to zero. We do not believe that this measurably affects any of our results, as the largest value changed in the above example was only negative eighteen.

Adapting the Massey model is the most difficult because of the specific interpretation that Massey attaches to his ratings. Subtracting the ratings of two teams is supposed to predict the point spread of a game between the two teams. If we load, for example, rushing yards into the $\mathbf{y}$ vector discussed previously and assign the $\mathbf{X}$ matrix with ones to teams that outgained their opponents in rushing yards and negative ones to teams that were outgained, then the ratings will subtract to give the predicted difference in rushing yards. This presents an interesting question, as outgaining an opponent in rushing yards does not necessarily correspond to victory. Instead of this interpretation, however, we consider the ratings to merely indicate the relative strength of the teams based on the given statistics. While the application of this model may be farther from the original intent of the author than our other adaptations, we believe it to be a valid interpretation, and, as we will see in Figure 1, it produces reasonable results.

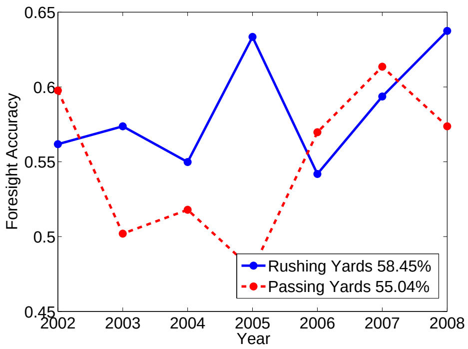
Figure 1 shows plots of foresight accuracy for each of the four models over seven seasons of NFL data, using rushing and passing yards as input. The average foresight accuracy for the seven year period is displayed in the legend.



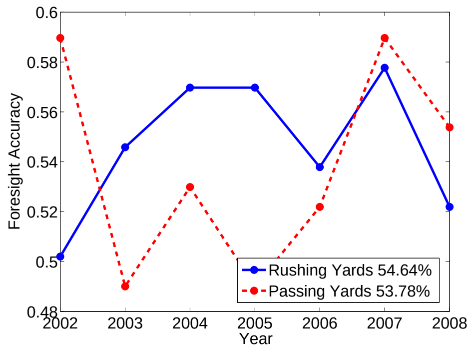
(a) Keener Model



(b) Massey Model



(c) Offense-Defense Model



(d) Generalized Markov Model

Figure 1: Rushing and Passing Yards in the Models

It is evident from the plots that rushing yards outperforms passing yards as a game predictor in all four models. In terms of average foresight accuracy for all seven years, rushing beats passing by 3.52% in the Keener model, 4.72% in the Massey model, 3.41% in the Offense-Defense model, 0.86% in the Generalized Markov model.

Since the Generalized Markov model is designed to use more than one statistic, we next run it on a convex combination of scores and rushing yards and a convex combination of scores and passing yards, respectively

$$\mathbf{G}_R = \alpha\mathbf{S} + (1-\alpha)\mathbf{R} \qquad \text{and} \qquad \mathbf{G}_P = \alpha\mathbf{S} + (1-\alpha)\mathbf{P}. \tag{15}$$

For each of the two inputs, we vary α from 0 to 1 to ensure that our choice of alpha does not arbitrarily affect the relative accuracies. Figure 2 is a plot of average foresight accuracy versus α for each. We denote the average foresight accuracy using $\mathbf{G}_R$ as A_R and the average foresight accuracy using $\mathbf{G}_P$ as A_P.

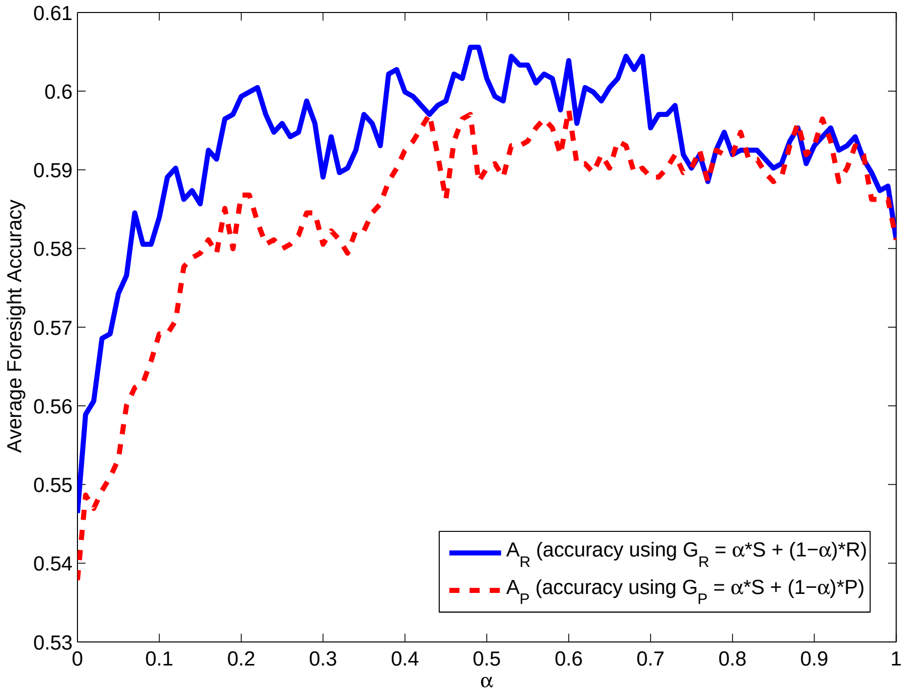


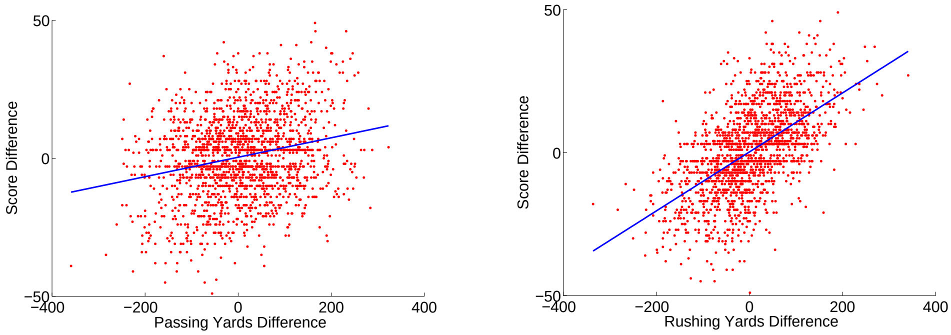
Figure 2: Scores with Rushing Yards and Passing Yards

Not surprisingly, A_R and A_P closely agree while $\alpha > 0.7$ since this is when a significant majority of the weight is on the score matrix. However, when $\alpha < 0.5$, $A_R > A_P$. The maximum for A_R is 60.56% at $\alpha = 0.49$ and the maximum for A_P is 59.76% at $\alpha = 0.61$ so that the maximum for A_R is

0.8% greater than the maximum of A_P. In this application of the Generalized Markov model it is again clear that rushing yards outperforms passing yards as a game predictor.

4.2 Correlation with Score Differences

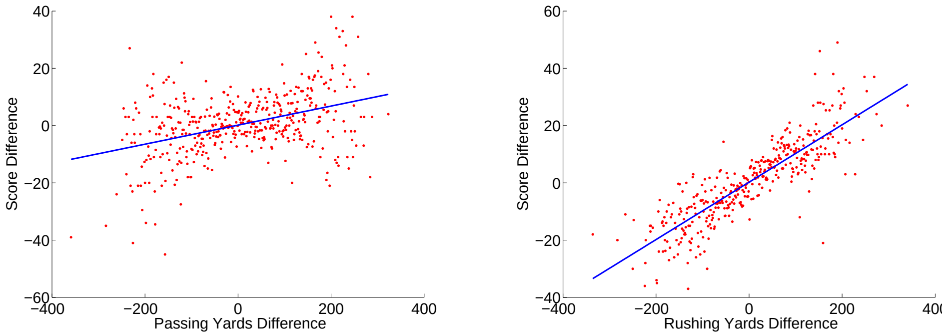
In this section we examine the relative correlation of differences between passing and rushing yards and differences in game scores. In the plots in Figure 3, each point represents one game, where the x-coordinate is the difference in either rushing or passing yards and the y-coordinate is the difference in game scores. If rushing or passing yards is truly important in determining the outcome of a game, then we expect to see that the more a team outgains another in yards, the more they will outscore them. Note that we do not need a strong correlation to draw a conclusion. We simply need one correlation to be stronger than the other.



(a) Score Difference v. Passing Difference (b) Score Difference v. Rushing Difference

Figure 3: Least-Squares Plots with all Data Points

It is visually obvious that the data for passing yards contain less of a trend than the data for rushing yards. The coefficients of determination (often denoted R^2) are 0.0569 for the passing yards data and 0.3333 for the rushing yards data. Thus, we can conclude that outgaining an opponent in terms of rushing yards is more correlated to outscoring that opponent than is outgaining them in terms of passing yards. As the data is compiled from seven NFL seasons (2002-2008), there are many games with the same yard difference, resulting in many "vertical lines" in our data. In Figure 4, we "clean" the data by plotting only the average score difference for repeated yard differences. We are motivated here by the desire to more clearly see the trend.



(a) Score Difference v. Passing Difference (b) Score Difference v. Rushing Difference

Figure 4: Least-Squares Plots with Averaged Data Points

4.3 Conditional Analysis

In order to analyze the relationship between rushing and passing yards from another viewpoint, we might consider certain conditional probabilities. For example, we might compare the conditional probability that a team outscores its opponent given that it outrushes its opponent with the conditional probability that a team outscores its opponent given that it outpasses its opponent. Such a comparison would shed new light on the correlation between outscoring and outrushing as compared to that between outscoring and outpassing. However, these probabilities would be very difficult to determine from the available data. Moreover, they would likely depend on many other factors, such as the strengths of the respective quarterbacks and the weather.

We can work around these issues by using what are, effectively, observational analogues of conditional probabilities. In particular, we will consider, for each season in the NFL, the fraction of games won by teams that did or did not outgain their opponents in rushing yards, passing yards, or total yards. These fractions are displayed in Figure 5.

It is apparent from Figure 5 that a team that outrushed its opponent is more likely to have won than a team that outpassed its opponent by at least 10% in every year. This suggests that outscoring is more strongly correlated with outrushing than with outpassing.

Note that the games that occurred in each season constitute a mere sample of a diverse population. As such, the fractions of games won under the given conditions are not necessarily the true probabilities. However, due to the consistency from year to year, it is reasonable to assume that the observed general pattern will continue to arise in subsequent years. In particular, observe that corresponding fractions are relatively constant over the last three years (2006-2008) as compared to the first four years. It seems as if the behavior of game outcome with respect to rushing and passing yards has in some sense stabilized in recent seasons. In contrast, the fractions of games won by teams that outgained in rushing yards, rushing yards but not passing yards, or passing but not rushing yards (respectively) are noticeably different in 2005 as compared to the other years; the 2005 season seems to be an outlier. A further investigation into these trends and the reasons behind them would be interesting, but is beyond the scope of this paper.

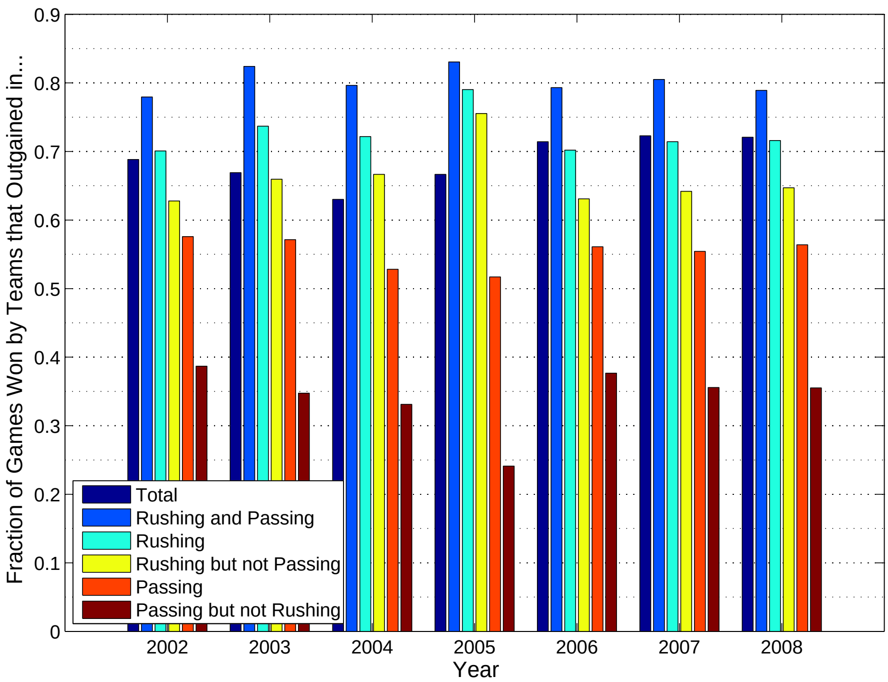


Figure 5: Fractions of games won by teams that outgained their opponents in rushing yards, passing yards, or total yards.

5 Interpretation of Results

At this point, it may be helpful to remind ourselves of the exact question we have been attempting to answer. We have sought to determine whether rushing or passing yards is a better game predictor and thus indicator of team strength. What we have *not* attempted to answer is whether rushing or passing is a better offensive strategy. This question has been proposed by Schatz (2005) and explored by Alamar (2006) and Rockerbie (2008). The difference is subtle but critical. For example, based on our results, a team might choose to exclusively rush the ball. After all, we found that when teams outrushed their opponent, they won more than 70% of the time. However, this is a distortion of our conclusions. The correct interpretation is that, in the course of a normally played game, outrushing one's opponent is more related to winning than is outpassing one's opponent.

We might call our viewpoint in this paper the "gambler's perspective," as opposed to the "coach's perspective." We have been attempting to identify statistical trends in the NFL that could allow us to better predict game outcomes. An intelligent gambler could look at our results and decide to base part of his betting decision on who he believes has a stronger rushing attack, which is a valid application of our conclusions. We are *not* taking the coach's perspective and attempting to tell teams how to play the game. A concerted effort to use our results as a coaching strategy would most likely change their usefulness as game predictors. The idea of the best offensive strategy is an interesting question, but we do not believe that our results answer it.

6 Conclusion

The central question of our paper has been whether rushing or passing yards is a better game predictor and thus superior indicator of team strength. Based on the methods that we used to investigate the question, the answer seems clear. Rushing yards is a better indicator of team strength than passing yards in every manner in which we compared them. In each of the four models we used, rushing yards clearly outperformed passing yards. Additionally, outrushing an opponent is more correlated with outscoring an opponent than is outpassing them. Finally, examining the historical trends associated with outgaining an opponent in a particular statistic clearly shows that in the past seven NFL seasons, teams that outgained their opponent in rushing yards won a higher percentage of games than teams that outgained their opponent in passing yards. Thus, if one wishes to predict the outcome of a particular NFL game and can gauge the relative strengths of the two team's rushing and passing games, one is more likely to choose the winner if one picks the team with the stronger rushing game. For anyone serious about the business of game prediction, more variables must be taken into account, but the relative strengths of the teams' rushing games are certainly important to consider.

References

Benjamin C. Alamar. The Passing Premium Puzzle. *Journal of Quantitative Analysis in Sports*, 2(4), 2006.

ESPN. NFL Expert Picks - Playoffs - Super Bowl, 2009. URL http://sports.espn.go.com/nfl/features/talent.

Anjela Y. Govan. *Ranking Theory with Application to Popular Sports*. PhD thesis, North Carolina State University, 2008.

Anjela Y. Govan, Amy N. Langville, and Carl D. Meyer. Offense-Defense Approach to Ranking Team Sports. *Journal of Quantitative Analysis in Sports*, 5(1), 2009.

James P. Keener. The Perron-Frobenius Theorem and the Ranking of Football Teams. *SIAM Review*, 1993.

Kenneth Massey. Statistical Models Applied to the Rating of Sports Teams. 1997.

Carl D. Meyer. *Matrix Analysis and Applied Linear Algebra*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2000.

Duane W. Rockerbie. The Passing Premium Puzzle Revisited. *Journal of Quantitative Analysis in Sports*, 4(2), 2008.

Aaron Schatz. Football's Hilbert Problems. *Journal of Quantitative Analysis in Sports*, 1(1), 2005.

Richard Sinkhorn and Paul Knopp. Concerning Nonnegative Matrices and Doubly Stochastic Matrices. *Pacific Journal of Mathematics*, 21(2):343–348, 1967.