THE DETERMINANTS OF SCORING IN 2010 NFL GAMES AND THE OVER/UNDER LINE

Tracy D. Rishel, Department of Management,
Susquehanna University, Selinsgrove, PA, rishel@susqu.edu

C. Barry Pfitzner and Steven D. Lang, Department of Economics/Business,
Randolph-Macon College, Ashland, VA, bpfitzne@rmc.edu, slang@rmc.edu

ABSTRACT

We collected data on a wide range of team statistics for the full 2010 National Football League season in order to compare estimates from 2010 to results obtained in prior research for the 2008 season. Based on regression analysis, we present evidence on the determinants of total points scored and the betting line for total points. The estimates for the 2010 season are similar to those obtained in 2008, especially for the betting line. Measures of how teams "match up," points scored in the immediately preceding game, and whether the game was played indoors play important roles in determining the betting line. A much smaller subset of variables are important in determining actual points scored. Most results are consistent across the two seasons.

INTRODUCTION

Many authors have investigated the market for betting on NFL games (see [1-8], for examples). Nearly all of these studies (except [7]) focus on the point spread, i.e., the difference between the points scored by the winning and losing teams.

In this paper we attempt to predict the total points scored and the over/under for National Football League (NFL) games played in the most recently completed NFL season, 2010-2011. We have collected and created variables that have been shown in prior research to be relevant to determining both the betting line and total points scored in a particular game. We estimate separate regression equations with the over/under line and actual points scored serving as dependent variables. These equations may be useful in confirming the results of research for earlier NFL seasons suggesting that variables measuring the ways teams "match up" against each other are important determinants of the "line" and total points scored.

DATA AND METHOD

The variables collected for this research include:

TP = total points scored for the home and visiting teams for each game played
PO = passing offense in yards per game
RO = rushing offense in yards per game
PD = passing defense in yards per game
RD = rushing defense in yards per game
TA = "takeaways" (turnovers gained per game)
GA = "giveaways" (turnovers lost per game)
D = a dummy variable equal to 1 if the game is played in a dome, 0 otherwise
PP = points scored by a given team in their prior game
L = the betting point spread (line) on the game

Match-ups Matter

The general regression format is based on the assumption that "match ups" are important in determining points scored in individual games. For example, if team "A" with the best passing offense is playing team "B" with the worst passing defense, ceteris paribus, team "A" would be expected to score many points. Similarly, a team with a very good rushing defense would be expected to allow relatively few points to a team with a poor rushing offense. In accord with this rationale, we formed the following variables:

PY = PO + PD = passing yards
RY = RO + RD = rushing yards

For example, suppose team "A" is averaging 325 yards (that's high) per game in passing offense and is playing team "B" which is giving up 330 yards (also, of course, high) per game in passing defense. The total of 655 would predict many passing yards will be gained by team "A," and likely many points will be scored by team "A." Subscripts are indicative of the visitor or home status of the variables, e.g., PY_v will represent the passing yard variable for the visiting team.

The dome variable will be a check to see if teams score more (or fewer) points if the game is played indoors. The variables on takeaways and giveaways will also be tested for relevance in scoring and setting the line.

The variable for points scored in the prior game (PP) is intended to check for streakiness in scoring. That is, if a team scores many (or few) points in a given game, are they likely to have a similar performance in the ensuing game?

The estimated equations may be useful in confirming (or contradicting) the results of the prior research, and may provide useful information applicable to wagering strategies.

RESULTS

All regressions reported in Table I are for individual games based on information known prior to the game. For example, if two teams are playing in week ten, only information known through week nine (rushing yards per game, passing yards per game, etc.) are utilized in the estimations for week ten.

The results of the regressions for 2008 and 2010 (we did not collect data for 2009) are contained in Table I. The second (2008) and fourth (2010) columns in the table are the regression with the line as the dependent variable. Every coefficient estimate for each year is correctly signed according to our expectations, statistically significant, and $\overline{R}^2 = .671$ for 2008 and $\overline{R}^2 = .665$ for 2010. Notice that the standard error is slightly lower for the 2010 season's betting lines, and the dome effect seems somewhat smaller for the 2010 line.

With the same set of explanatory variables, the third and fifth columns in Table I contain the results for actual points scored in the games. While regressions for the line explain fully two-thirds of the variance in that dependent variable, the equations for the actual points explains only 5.2 percent of the variance in total points for 2008 and only 5.7 percent for 2010. Further, only four of the seven explanatory variables meeting the test for statistical significance at traditional levels for 2008, and only three for 2010. The F-test for overall significance of the equation for total points does indicate, however, that a significant portion of the variance in the dependent variable is explained by the regression equation for 2008, but not for 2010. More parsimonious specifications of the 2010 regression for points scored does indicate overall statistical significance. For these regressions on total points, notice there is no dome effect for the 2010 regression. The line is, as expected, much easier to predict than actual points scored. That is, the outcomes of the games and points scored are not easily predicted, which is "why they play the games."

At least two further observations are in order. First, consider the coefficients for points scored in the previous game. Those variables matter in determining the line for the game for both the 2008 (PP_h would pass a test of significance at the 90% level of confidence) and 2010 seasons. However, they seem to play an insignificant (statistical or practical) role in determining the actual points scored. This particular result may mean that bettors place too much emphasis on recent information, as other authors have suggested. Second, for the 2008 results, we tentatively suggested that the "line" underestimated the dome effect as compared to the actual points scored (the coefficient is larger for actual points equation in 2008). That conclusion is not supported by the results for the 2010 NFL season.

Table I: Regression Results for the Line and Total Points, 2008 and 2010 Seasons

Explanatory Variable	2008 Dependent Variable = Line	2008 Dependent Variable = Total Points	2010 Dependent Variable = Line	2010 Dependent Variable = Total Points
Intercept	-21.03 (-5.29)	-10.58 (-0.59)	-5.79 (-1.90)	-16.09 (-0.75)
PY_h	0.0476* (12.10)	0.0166 (0.94)	0.0380* (12.74)	0.042** (1.98)
RY_h	0.0507* (6.87)	0.0559** (1.69)	0.0310* (6.22)	0.038 (1.08)
PY_v	0.0442* (11.52)	0.0376** (2.18)	0.0378* (12.47)	0.038** (1.79
RY_v	0.0450* (5.93)	0.0576** (1.69)	0.0253* (4.94)	0.088* (2.44)
PP_v	0.0669* (2.86)	0.0766 (0.73)	0.0531* (3.61)	-0.029 (-0.276)
PP_h	0.0343*** (1.53)	0.1100 (1.09)	0.0348* (2.38)	-0.077 (-0.744)
D	2.21* (4.01)	5.17** (2.09)	0.99* (2.86)	-0.096 (-0.039)
$\bar{R}^2$	0.671	0.052	0.665	0.057
SEE	2.84	12.76	2.012	14.13
N	194	194	192	192
F-stat	57.2*	2.51**	52.22*	1.588

(The numbers in parentheses are t-statistics)
*** represents significance at the 90 percent level of confidence or better,
** represents significance at the 95 percent level of confidence or better, and
* represents significance at the 99 percent level of confidence or better for one-tailed tests.

The variables for turnovers (giveaways and takeaways) played no statistical role in predicting either the line or total points. The number of (expected) turnovers in a particular game is an ambiguous determinant of points scored in a particular game, and we found no statistical role in any of the 2010 regressions.

CONCLUSIONS

Employing explanatory variables known prior to game time, this research presents successful equations for predicting the betting line on individual NFL games. Across two seasons, we are able to explain two-thirds of the variance in the betting lines for individual games. Importantly, we make no allowance for injuries, weather in outdoor games, or any other variables of that sort. The regressions for the line are very consistent across the two years, with the exception of a reduced estimate of the effect of playing "indoors." The regressions for actual points scored, an admittedly difficult variable to predict, explain a much smaller portion of the variance of that variable.

The data collected for the 2010 season will be employed in a future project to predict scoring for the visiting and home teams separately, and attempting to use those predicted totals to test the efficient markets hypothesis for the wagering lines for NFL games.

REFERENCES

[1] Borghesi, Richard, "The Home Team Weather Advantage and Biases in the NFL Betting Market," *Journal of Economics and Business*, 2007, 59, 340-354.

[2] Dare, William H. and S.S. MacDonald, "A Generalized Model for Testing Home and Favourite Team Advantage in Point Spread Markets," *Journal of Financial Economics*, 40, 1996, 295-318.

[3] Dare, William H. and A. Steven Holland, "Efficiency in the NFL Betting Market: Modifying and Consolidating Research Methods," *Applied Economics*, 2004, 36, 9-15.

[4] Gray, Philip K. and Stephen F. Gray, "Testing Market Efficiency: Evidence from the NFL Sports Betting Market," *The Journal of Finance*, 1997, LII(4), 1725-1737.

[5] Levitt, Steven D., "How Do Markets Function? An Empirical Analysis of Gambling on the National Football League," *National Bureau of Economic Research working paper* 9422, 2002.

[6] Paul, R.J. & Weinbach, A.P. "Market Efficiency and a Profitable Betting Rule: Evidence from Totals on Professional Football," *Journal of Sports Economics*, 2002, 3, 256-263.

[7] Pfitzner, C. Barry, Steven D. Lang and Tracy D. Rishel, "The Determinants of Scoring in NFL Games and Beating the Over/Under Line," *New York Economic Review*, Volume 40, Fall 2009.

[8] Vergin, Roger C., "Overreaction in the NFL Point Spread Market," *Applied Financial Economics*, 2001, 11, 497-509.